



(51) International Patent Classification:  
**G06F 17/30** (2006.01) **G06F 17/40** (2006.01)

(21) International Application Number:  
PCT/US2010/031090

(22) International Filing Date:  
14 April 2010 (14.04.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
61/169,662 15 April 2009 (15.04.2009) US

(71) Applicant (for all designated States except US): **EVRI INC.** [US/US]; 71 Columbia St., Suite 300, Seattle, WA 98104 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **WISSNER, James, M.** [US/US]; 1433 Battery Caulfield Road, San Francisco, CA 94129 (US). **SPIVACK, Nova, T.** [US/US]; 547 29th Street, #7, San Francisco, CA 95407 (US).

(74) Agents: **FU, Yenyun** et al.; Perkins Coie LLP, P.O. Box 1208, Seattle, WA 98111-1208 (US).

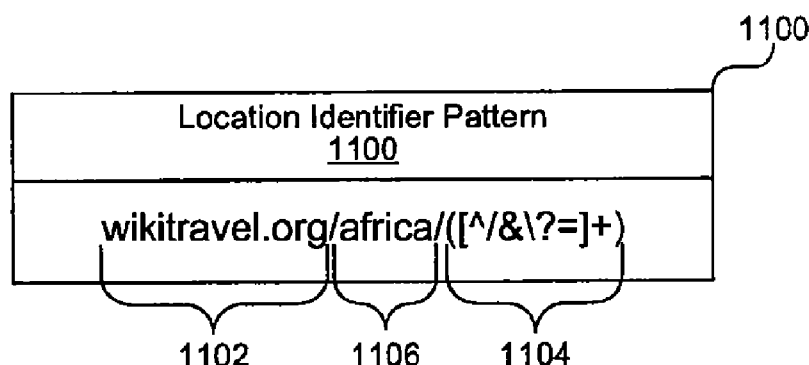
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG,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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: SEARCH AND SEARCH OPTIMIZATION USING A PATTERN OF A LOCATION IDENTIFIER



**FIG. 11A**

(57) Abstract: Systems and methods for search and search optimization using a pattern in a location identifier is disclosed. In one aspect, embodiments of the present disclosure include a method, which may be implemented on a system, of search and search optimization. The method includes, detecting a set of location identifiers that have a pattern that matches a specified pattern and identifying a set of search results as having content related to the semantic type. The specified pattern can be stored in a computer-readable storage medium and corresponds to a semantic type. The set of search results can include objects associated with the set of location identifiers having the specified pattern.

## SEARCH AND SEARCH OPTIMIZATION USING A PATTERN OF A LOCATION IDENTIFIER

### CROSS-REFERENCE TO RELATED APPLICATIONS

- 5 [0001] This application claims priority to U.S. Provisional Patent Application No. 61/169,662 filed 15 April 2009 which is hereby incorporated by reference in its entirety.

### TECHNICAL FIELD

- [0002] This technology relates generally to search and search optimization of objects containing electronic content and in particular, to optimizing search by enhancing  
10 relevancy through semantics and by improving computation time for the semantics enhancement.

### BACKGROUND

- [0003] The search through the vast amount of resources containing electronic content (e.g., digital music, digital video, documents, text files, web pages) in the digital world is  
15 increasingly becoming a resource consuming task. The mere task of executing the search based on a basic keyword comparison is daunting regardless of whether the scope of the search is within the confines of a local computing system, a private network, a local area network, or the World Wide Web.

- [0004] The task of producing meaningful search results or even customized search results  
20 poses a yet further significant and seemingly impossible barrier to improving or enhancing a user's search experience. Furthermore, the prevalence of user-added or user-supplemented content has added to the difficulty and intricacy of performing search in an expedited manner while simultaneously producing meaningful results among the vast amount of available locally stored, shared, or globally available resources.

**BRIEF DESCRIPTION OF THE DRAWINGS**

[0005] FIG. 1 illustrates an example block diagram of a host server that performs techniques of search and search optimization of content from various content providers to users of client devices through a network.

5 [0006] FIG. 2 depicts an example block diagram of the components of a host server that performs techniques of search and search optimization.

[0007] FIG. 3A depicts an example functional block diagram of the host server that performs search and optimizes the search process.

10 [0008] FIG. 3B depicts an example block diagram illustrating the components of the host server that performs search and/or optimized search.

[0009] FIG. 4 depicts a flow chart illustrating an example process for optimizing search by identifying web pages having content of or related to a semantic type.

15 [0010] FIG. 5 depicts a flow chart illustrating an example process for optimizing search by identifying web pages having content associated with multiple attributes of the semantic type.

[0011] FIG. 6 depicts a flow chart illustrating an example process for search of objects having content of or related to a semantic type using patterns of the location identifiers of the objects.

20 [0012] FIG. 7 depicts a flow chart illustrating an example process for optimizing search using semantics.

[0013] FIG. 8 depicts a flow chart illustrating an example process for generating user-customized search results using user-defined semantic types.

[0014] FIG. 9 depicts a flow chart illustrating an example process for building a semantics-enhanced search engine.

25 [0015] FIG. 10 depicts a flow chart illustrating an example process for performing a search by mapping a search query into a specified pattern in a pattern template.

[0016] FIG. 11A-B illustrate examples of location identifier patterns.

[0017] **FIG. 12A** illustrates an example screenshot showing examples of location identifier patterns and associated semantic types.

[0018] **FIG. 12B** illustrates an example screenshot of a location identifier and various editable fields.

5 [0019] **FIG. 12C** illustrates an example screenshot of another location identifier and various editable fields.

[0020] **FIG. 13A** illustrates another example screenshot showing examples of location identifier patterns and an associated semantic type for the domain.

10 [0021] **FIG. 13B** illustrates an example screenshot of a location identifier and various editable fields.

[0022] **FIG. 13C** illustrates another example screenshot of a location identifier and various editable fields.

[0023] **FIG. 14A** illustrates an example template for specifying how a search result having content from a specific domain of a particular semantic type is displayed.

15 [0024] **FIG. 14B** illustrates an example stylesheet for specifying how a search result having content from a specific domain of a particular semantic type is displayed.

[0025] **FIG. 15** illustrates an example list of domains for which location identifier patterns have been extracted.

20 [0026] **FIG. 16A** illustrates an example list of vocabulary showing semantic types that have been defined.

[0027] **FIG. 16B** illustrates an example list of vocabulary showing attributes of semantic types that have been defined.

25 [0028] **FIG. 17** illustrates an example screenshot of a user interface of a search engine that executes search and/or optimizes search by performing pattern recognition using location identifier patterns.

[0029] **FIG. 18** illustrates a screenshot showing an example list of search results identified using keyword search without further semantic-categorization.

[0030] **FIG. 19A** illustrates a screenshot showing an example list of search results identified responsive to a search query and categorized as having content related to a semantic type.

5 [0031] **FIG. 19B** illustrates a screenshot showing an example of a search result hosted by a sponsored domain and identified as having content related to a semantic type.

[0032] **FIG. 20** illustrates a screenshot showing an example list of search results identified responsive to the search query and categorized as a news article.

10 [0033] **FIG. 21** illustrates a screenshot showing an example list of search results identified responsive to the search query and categorized as having content related to another semantic type.

[0034] **FIG. 22** illustrates a screenshot showing an example list of search results identified responsive to the search query and categorized as having content related to yet another semantic type.

15 [0035] **FIG. 23** illustrates an example list of search results identified responsive to the search query and categorized as having content related to yet another semantic type.

[0036] **FIG. 24** illustrates an example of a web-dialogue that is parsed to obtain information regarding a particular webpage.

[0037] **FIG. 25** depicts a flow chart illustrating an example process for semantic search enhanced advertising.

20 [0038] **FIG. 26** depicts a flow chart illustrating an example process for targeting advertisements using semantic constraints.

[0039] **FIG. 27** shows a diagrammatic representation of a machine in the example form of a computer system within which a set of instructions, for causing the machine to perform any one or more of the methodologies discussed herein, may be executed.

**DETAILED DESCRIPTION**

[0040] The following description and drawings are illustrative and are not to be construed as limiting. Numerous specific details are described to provide a thorough understanding of the disclosure. However, in certain instances, well-known or conventional details are not described in order to avoid obscuring the description. References to one or an embodiment in the present disclosure can be, but not necessarily are, references to the same embodiment; and, such references mean at least one of the embodiments.

[0041] Reference in this specification to “one embodiment” or “an embodiment” means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the disclosure. The appearances of the phrase “in one embodiment” in various places in the specification are not necessarily all referring to the same embodiment, nor are separate or alternative embodiments mutually exclusive of other embodiments. Moreover, various features are described which may be exhibited by some embodiments and not by others. Similarly, various requirements are described which may be requirements for some embodiments but not other embodiments.

[0042] The terms used in this specification generally have their ordinary meanings in the art, within the context of the disclosure, and in the specific context where each term is used. Certain terms that are used to describe the disclosure are discussed below, or elsewhere in the specification, to provide additional guidance to the practitioner regarding the description of the disclosure. For convenience, certain terms may be highlighted, for example using italics and/or quotation marks. The use of highlighting has no influence on the scope and meaning of a term; the scope and meaning of a term is the same, in the same context, whether or not it is highlighted. It will be appreciated that same thing can be said in more than one way.

[0043] Consequently, alternative language and synonyms may be used for any one or more of the terms discussed herein, nor is any special significance to be placed upon whether or not a term is elaborated or discussed herein. Synonyms for certain terms are provided. A recital of one or more synonyms does not exclude the use of other synonyms. The use of examples anywhere in this specification including examples of any terms discussed herein is illustrative only, and is not intended to further limit the scope and meaning of the

disclosure or of any exemplified term. Likewise, the disclosure is not limited to various embodiments given in this specification.

[0044] Without intent to further limit the scope of the disclosure, examples of instruments, apparatus, methods and their related results according to the embodiments of the present disclosure are given below. Note that titles or subtitles may be used in the examples for convenience of a reader, which in no way should limit the scope of the disclosure. Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. In the case of conflict, the present document, including definitions will control.

10 [0045] Embodiments of the present disclosure include systems and methods for performing search and search optimization using a pattern of a location identifier.

[0046] The techniques described herein generally relate to the search and the search optimization of electronic content including but not limited to electronic content that is distributed throughout the Internet or World Wide Web. The location identifier of electronic content or a source can be a web address that a web browser uses to locate the content of the source for access by a user via a web browser (e.g., Universal Resource Identifier (URI) or Universal Resource Locator (URL)).

15 [0047] In one example of the techniques described herein as applied to search for online content distributed via the Internet, information derived directly from extracting patterns and words in URLs and/or URIs provide a wealth of semantic information. The semantic information derived from the URLs and/or URIs can be used in searching for content based on semantics and/or for semantically optimizing the search process.

20 [0048] By using URL/URI patterns for online content, the semantic information can be deduced and extracted using relatively lesser computing resources, effort, and time, for example, as compared to solely extracting such semantic information by performing various data mining techniques on the content source. Pattern recognition using URL/URI or other location identifiers allows semantic information of a vast number of content sources to be determined in a relatively short amount of time compared to traditional mechanism. The semantic information thus extracted facilitates and expedites the search process and to generate semantically-valuable search results.

30

[0049] In other words, the patterns extracted from URLs/URIs can be correlated to ontologically defined semantic types and their attributes (e.g., properties). In addition, patterns can be drawn between non-semantically marked up HTML or XHTML content and semantic types/attributes. In addition to using the correlated patterns and the extracted semantic information for enhanced search engine indexing, the patterns (e.g., from URIs/URLs and/or (X)HTML content) can be associated directly to types, properties, and/or attributes defined in one or more ontologies.

[0050] For example, metadata (e.g., Resource Description Framework (RDF) triples) can be automatically extracted from XML or XHTML documents (e.g., using GRDDL (Gleaning Resource Descriptions From Dialects of Languages)) to convert (X)HTML or other content to semantic content. This conversion and extraction of semantic data from content can be used for building or enhancement of a search engine.

[0051] Note that the processes and techniques herein described are not limited to search and search optimization of web content or other online/shared content but also for search and search optimization of content sources stored locally on one computer/server, on a network of computers, and/or shared over a private network. Location identifiers and patterns thereof that can be used for search optimization, although described generally as including URIs and URLs, are not limited to such. Location identifiers such as file directories, file paths, network locations, and/or any other form of location identifiers for electronic content sources and the patterns thereof are contemplated and are also considered to be within the novel art of the disclosure.

[0052] FIG. 1 illustrates an example block diagram of a host server 100 that performs techniques of search and search optimization of content from various content providers 108A-N to users of client devices 102A-N through a network 106.

[0053] The client devices 102A-N can be any system and/or device, and/or any combination of devices/systems that is able to establish a connection with another device, a server and/or other systems. Client devices 102A-N each typically include a display and/or other output functionalities to present information and data exchanged between among the devices 102A-N and the host server 100. For example, the client devices 102A-N can be any of, but are not limited to, a server desktop, a desktop computer, a computer cluster, or portable devices including, a notebook, a laptop computer, a handheld computer, a palmtop



computer, a mobile phone, a cell phone, a smart phone, a PDA, a Blackberry device, a Treo, and/or an iPhone, etc. In one embodiment, the client devices 102A-N and content providers 108A-N of electronic content are coupled to a network 106. In some embodiments, the devices 102A-N and host server 100 may be directly connected to one another.

[0054] In one embodiment, the host server 100 is operable to perform search and/or provide search optimization techniques and related services. The search and search optimization can be performed using patterns of location identifiers (e.g., web addresses, URIs, URLs, file names, directories, network addresses, pathways, etc.) of content sources (e.g., sources of electronic content).

[0055] The search optimization performed by the host server 100 can include semantics-enhanced search that identifies and/or further categorizes web pages or other objects as having content of or being related specific semantic types. Further, the host server 100 can further categorize/identify objects as having content or content related to certain attributes of semantic types. In one embodiment, the host server 100 generates user-customized search results using user-defined semantic types.

[0056] Functions and techniques performed by the host server 100 and the components therein are described in detail with further references to the examples of FIG. 3A-3B.

[0057] The client devices 102A-N are generally operable to provide user access (e.g., visible access, audible access) to the search process, the user-customized search process, and the optimized search (e.g., via semantics), for example via user interface 104A-N displayed on the display units. The client devices 102A-N may also be used (e.g., by programmers and developers) for building a semantics-enhanced search engine or by end users to customize their search results, for example.

[0058] The network 106, over which the client devices 102A-N and the host server 100 communicate, may be a telephonic network, an open network, such as the Internet, or a private network, such as an intranet and/or the extranet. For example, the Internet can provide file transfer, remote log in, email, news, RSS, and other services through any known or convenient protocol, such as, but is not limited to the TCP/IP protocol, Open System Interconnections (OSI), FTP, UPnP, iSCSI, NSF, ISDN, PDH, RS-232, SDH, SONET, etc.

**[0059]** The network 106 can be any collection of distinct networks operating wholly or partially in conjunction to provide connectivity to the client devices 102A-N and the host server 100 and may appear as one or more networks to the serviced systems and devices. In one embodiment, communications to and from the client devices 102A-N can be  
5 achieved by, an open network, such as the Internet, or a private network, such as an intranet and/or the extranet. In one embodiment, communications can be achieved by a secure communications protocol, such as secure sockets layer (SSL), or transport layer security (TLS).

**[0060]** In addition, communications can be achieved via one or more wireless networks,  
10 such as, but are not limited to, one or more of a Local Area Network (LAN), Wireless Local Area Network (WLAN), a Personal area network (PAN), a Campus area network (CAN), a Metropolitan area network (MAN), a Wide area network (WAN), a Wireless wide area network (WWAN), Global System for Mobile Communications (GSM),  
Personal Communications Service (PCS), Digital Advanced Mobile Phone Service (D-  
15 Amps), Bluetooth, Wi-Fi, Fixed Wireless Data, 2G, 2.5G, 3G networks, enhanced data rates for GSM evolution (EDGE), General packet radio service (GPRS), enhanced GPRS, messaging protocols such as, TCP/IP, SMS, MMS, extensible messaging and presence protocol (XMPP), real time messaging protocol (RTMP), instant messaging and presence  
protocol (IMPP), instant messaging, USSD, IRC, or any other wireless data networks or  
20 messaging protocols.

**[0061]** The host server 100 may include internally or be externally coupled to a user repository 128, a knowledge repository 130, and/or a user content repository 132. The repositories can store software, descriptive data, images, system information, drivers, and/or any other data item utilized by other components of the host server 100 and/or any  
25 other servers for operation. The repositories may be managed by a database management system (DBMS), for example but not limited to, Oracle, DB2, Microsoft Access, Microsoft SQL Server, PostgreSQL, MySQL, FileMaker, etc.

**[0062]** The repositories can be implemented via object-oriented technology and/or via text files, and can be managed by a distributed database management system, an object-  
30 oriented database management system (OODBMS) (e.g., ConceptBase, FastDB Main Memory Database Management System, JDOInstruments, ObjectDB, etc.), an object-relational database management system (ORDBMS) (e.g., Informix, OpenLink Virtuoso,

VMDS, etc.), a file system, and/or any other convenient or known database management package.

[0063] In some embodiments, the host server 100 is able to provide data to be stored in the user repository 128, the knowledge repository 130, and the user content repository 132

5 and/or can retrieve data stored in the user repository 128, the knowledge repository 130, and/or the user content repository 132. The user repository 128 can store user information, user preferences, user search preferences, sharing preferences, user-defined semantic types, user-specified attributes of semantic types, user-specified patterns for location identifiers, user reviews of electronic content and sources, device information, hardware information,  
10 etc.

[0064] The knowledge repository 130 can include, by way of example but not limitation, sets of ontologies, taxonomies and/or folksonomies that can be used by the host server 100 for determining the semantic type or attribute type of electronic content provided in online sources to enhance or optimize search. The ontologies and taxonomies may be user-

15 defined and used on a per-user basis for enhancing search. The ontologies and taxonomies may also be provided by another source or centrally defined and applied globally or to users who have not defined their own ontologies. The knowledge repository 130 in some instances, can also include, dictionaries, thesauruses, and/or encyclopedias, etc. In one embodiment, location identifier patterns and their associated semantic types and/or  
20 attributes are stored in the knowledge repository 130.

[0065] The user content repository 132 can include, user content including, by way of example but not limitation, user-generated, user-edited, and/or user-provided electronic content (e.g., documents, articles, audio files, video files, text files, images, user dialogue, user chat, etc.). The user content can be aggregated from a host or multiple hosts of online  
25 locations where users share information including but not limited to various networking sites (e.g., blogs, forums, social messaging site, social networking site, business networking, knowledge networking, etc.).

[0066] Repository 132 can also include user-supplemented content including content and edits, reviews, and/or comments provided by users. The host server 100 can perform  
30 search and search optimization not only on web content but also on user-generated, user-

provided, and/or user supplemented content contributed through single or multiple content sharing or networking sites.

[0067] FIG. 2 depicts an example block diagram of the components of a host server 200 that performs techniques of search and search optimization.

5 [0068] In the example of FIG. 2, the host server 200 includes a network controller 202, a firewall 204, a multimedia server 206, an application server 208, a web application server 212, a search engine 214, and a database including a database storage 216 and database software 218.

10 [0069] In the example of FIG. 2, the network controller 202 can be a networking device that enables the host server 100 to mediate data in a network with an entity that is external to the host server 100, through any known and/or convenient communications protocol supported by the host and the external entity. The network controller 202 can include one or more of a network adaptor card, a wireless network interface card, a router, an access point, a wireless router, a switch, a multilayer switch, a protocol converter, a gateway, a  
15 bridge, a bridge router, a hub, a digital media receiver, and/or a repeater.

[0070] The firewall 204, can, in some embodiments, govern and/or manage permission to access/proxy data in a computer network, and track varying levels of trust between different machines and/or applications. The firewall 204 can be any number of modules having any combination of hardware and/or software components able to enforce a  
20 predetermined set of access rights between a particular set of machines and applications, machines and machines, and/or applications and applications, for example, to regulate the flow of traffic and resource sharing between these varying entities. The firewall 204 may additionally manage and/or have access to an access control list which details permissions including for example, the access and operation rights of an object by an individual, a  
25 machine, and/or an application, and the circumstances under which the permission rights stand.

[0071] Other network security functions can be performed or included in the functions of the firewall 204, can be, for example, but are not limited to, intrusion-prevention, intrusion detection, next-generation firewall, personal firewall, etc. without deviating from the novel  
30 art of this disclosure. In some embodiments, the functionalities of the network controller

202 and the firewall 204 are partially or wholly combined and the functions of which can be implemented in any combination of software and/or hardware, in part or in whole.

[0072] In the example of **FIG. 2**, the host server 200 includes the multimedia server 206 or a combination of multimedia servers to manage images, photographs, animation, video, audio content, graphical content, documents, and/or other types of multimedia data for use in or to supplement search results or semantics-enhanced search results. The multimedia server 206 is any software suitable for delivering messages to facilitate retrieval/transmission of multimedia data among servers to be provided to other components and/or systems of the host server 200, for example while rendering a web page or other types of objects (e.g., electronic content sources) including multimedia content.

[0073] In addition, the multimedia server 206 can facilitate transmission/receipt of streaming data such as streaming images, audio, and/or video. The multimedia server 206 can be configured separately or together with the web application server 212, depending on a desired scalability of the host server 200. Examples of graphics file formats that can be managed by the multimedia server 206 include but are not limited to, ADRG, ADRI, AI, GIF, IMA, GS, JPG, JP2, PNG, PSD, PSP, TIFF, and/or BMP, etc.

[0074] The application server 208 can be any combination of software agents and/or hardware modules for providing software applications to end users, external systems and/or devices. For example, the application server 208 provides specialized or generic software applications that provides search or search optimization functionalities which may be semantics-enhanced. The software applications provided by the application server 208 can be automatically uploaded/downloaded on-demand on an as-needed basis or manually at the user's request.

[0075] The software applications provided by the application server 208, for example, can allow end users to define semantic types and/or to customize the search experience. The software applications provided by the application server, can, in one embodiment, allow programmers or developers to build semantics-enhanced search engines using pattern recognition of location identifiers (e.g., URLs and/or URIs). The application server 208 can also facilitate interaction and communication with the web application server 212, or with other related applications and/or systems. The application server 208 can in some

instances, be wholly or partially functionally integrated with the web application server 212.

[0076] The web application server 212 can include any combination of software agents and/or hardware modules for accepting Hypertext Transfer Protocol (HTTP) requests from end users, external systems, and/or external client devices and responding to the request by providing the requestors with web pages, such as HTML documents and objects that can include static and/or dynamic content (e.g., via one or more supported interfaces, such as the Common Gateway Interface (CGI), Simple CGI (SCGI), PHP, JavaServer Pages (JSP), Active Server Pages (ASP), ASP.NET, etc.).

[0077] In addition, a secure connection, SSL and/or TLS can be established by the web application server 212. In some embodiments, the web application server 212 renders the user interfaces in the example screenshots of **FIG. 12-24**. The user interfaces provided by the web application server 212 to client users/end devices provide the user interface screens 104A-104N for example, to be displayed on client devices 102A-102N in the example of **FIG. 1**.

[0078] In one embodiment, the host server 200 includes a search engine 214 for performing search and search optimization, for example, using patterns in location identifiers of objects such as web pages or other types of content. The objects can include electronic content including but not limited to, electronic documents, digital audio, digital video, images, graphical content, interactive content, etc. The search engine 214 can optimize the search process by enhancing relevancy of search results through, for example, semantics. The search engine can also optimize search by expediting the process to optimize the quantity of sources that can be analyzed in a given time period with limited computing power.

[0079] The databases 216, 218 can store software, descriptive data, images, system information, drivers, and/or any other data item utilized by other components of the host server 200 for operation. The databases 216, 218 may be managed by a database management system (DBMS), for example but not limited to, Oracle, DB2, Microsoft Access, Microsoft SQL Server, PostgreSQL, MySQL, FileMaker, etc. The databases 216, 218 can be implemented via object-oriented technology and/or via text files, and can be managed by a distributed database management system, an object-oriented database

management system (OODBMS) (e.g., ConceptBase, FastDB Main Memory Database Management System, JDOInstruments, ObjectDB, etc.), an object-relational database management system (ORDBMS) (e.g., Informix, OpenLink Virtuoso, VMDS, etc.), a file system, and/or any other convenient or known database management package.

5 [0080] In the example of **FIG. 2**, the host server 200 includes components (e.g., a network controller, a firewall, a storage server, an application server, a web application server, a search engine, and/or a database including a database storage and database software, etc.) coupled to one another and each component is illustrated as being individual and distinct.

[0081] However, in some embodiments, some or all of the components, and/or the  
10 functions represented by each of the components can be combined in any convenient or known manner. Furthermore, the functions represented by the devices can be implemented individually or in any combination thereof, in hardware components, software agents, or a combination of hardware components and software agents.

[0082] **FIG. 3A** depicts an example functional block diagram of the host server 300 that  
15 performs search and optimizes the search process.

[0083] The host server 300 includes a network interface 302, a location identifier retrieval module 304, an extraction module 306, a user module 308, a knowledge management module 310, a user interface module 312, and a search engine 314. The search engine 314 includes a recognition module 316, a results generator module 318, and a search optimizer  
20 module 320. In one embodiment, the location identifier retrieval module includes a URI identification module 305. The host server 300 may further include a data mining module 322, a dialogue analyzer module 324, a sponsorship module 326, and/or an advertisement module 328.

[0084] In one embodiment, the host server 300 is coupled to a user repository 328, a  
25 knowledge repository 330, and/or a user content repository 332. The user repository 328, knowledge repository 330, and the user content repository 332 have been described with further reference to the example of **FIG. 1**.

[0085] Additional or less modules can be included without deviating from the novel art of this disclosure. In addition, each module in the example of **FIG. 3A** can include any

number and combination of sub-modules, and systems, implemented with any combination of hardware and/or software modules.

[0086] The host server 300, although illustrated as comprised of distributed components (physically distributed and/or functionally distributed), could be implemented as a collective element. In some embodiments, some or all of the modules, and/or the functions represented by each of the modules can be combined in any convenient or known manner. Furthermore, the functions represented by the modules can be implemented individually or in any combination thereof, partially or wholly, in hardware, software, or a combination of hardware and software.

10 [0087] In the example of **FIG. 3A**, the network interface 302 can be a networking device that enables the host server 300 to mediate data in a network with an entity that is external to the host server, through any known and/or convenient communications protocol supported by the host and the external entity. The network interface 302 can include one or more of a network adaptor card, a wireless network interface card, a router, an access  
15 point, a wireless router, a switch, a multilayer switch, a protocol converter, a gateway, a bridge, a bridge router, a hub, a digital media receiver, and/or a repeater.

[0088] One embodiment of the host server 300 includes a location identifier retrieval module 304. The location identifier retrieval module 304 can be any combination of software agents and/or hardware modules able to detect, identify, and/or retrieve a location  
20 identifier of an object having electronic content.

[0089] An object can be any electronic object stored, shared, distributed, and/or accessed through or using computing devices including but not limited to electronic documents, articles, audio files, video files, multimedia content, interactive content, emails, a web page, a link, an electronic review, a dialogue, a file, a directory, an online message, etc.  
25 The location identifier of the object can be a location identifier in the World Wide Web, a private network, a local network, or local storage on computing element/device including but not limited to a web address (e.g., a URI and/or a URL), file directories, file paths, a network address, a file location, etc.

[0090] For a web page, the location identifier retrieval module 304 can identify a location  
30 identifier (e.g., URI or URL) of the web page (e.g., by the URI identification module of the location identifier retrieval module 304). In one embodiment, location identifiers of web



pages in a domain are identified from the domain's sitemap. In other embodiments, URIs and URLs can be manually identified or seeded from other resources. For example, a web crawler (e.g., a web crawl corpus such as Build Your Own Search Service (BOSS) platform) can be used to filter domains. In addition, a manual crawl of the site can be performed to extract URLs and/or URIs in domains. Other known and/or convenient methods can also be used to extract location identifiers. For other types of objects, the associated types of location identifiers can be identified and retrieved by the location identifier retrieval module 304.

[0091] Note that semantic types of content in objects (e.g., web pages) can be partially or fully automatically determined by the system or specified by an end user. For example, semantic type can be automatically determined through topic detection, natural language processing (NLP), speech processing, latent semantics indexing, etc. Semantic types can also be defined by the end user through tagging or annotating the object (e.g., web page) through a user interface in which the object is provided for access.

[0092] Note that each semantic type can be associated with one or more attributes that may be user-defined, automatically determined by the system, or a combination of both. Attributes of semantic types may be determined through a predefined ontology or user defined ontology. For example, a semantic type of 'Restaurant' can include the attributes 'Location', 'Price Range', 'Cuisine', etc. Some of these attributes may be system defined and some may be user-specified. Attributes defined by users may be provided to other users for reference and/or for rating, for example, via a user interface.

[0093] One embodiment of the host server 300 includes an extraction module 306. The extraction module 306 can be any combination of software agents and/or hardware modules able to extract a pattern from a location identifier of an object.

[0094] For example, the extraction module 306 can extract a pattern from the URI of a web page. In one embodiment, the extracted pattern corresponds to the semantic type of the content of the web page and can be used, for example, by the search engine 314 of the host server 300 to identify other objects (e.g., web pages) having content of or related to the same semantic type. The pattern of the location identifier once identified, can be stored in a computer-readable storage medium (e.g., in a computer database stored in hardware storage) for subsequent usage when determining the semantic type of content in an object

(e.g., web page, off-line document, chat dialogue, etc.) to enhance the relevancy of search results and/or to identify and generate search results having content of or related to the semantic type.

[0095] In one embodiment, the extraction module 306, based on the semantic type of the content in the associated object or web page, automatically parses the location identifier to extract the pattern based on the semantic type. For example, the location identifier may include one or more portions with words that correspond to the semantic type in literal translation (e.g., in meaning). These portions may be defined in the pattern of the location identifier as the semantic type segment. For example, a semantic type of "car reviews" may correspond to a web page having a portion with the term "car" or "car/review". The term "car" or "car/review" may then be automatically identified and used in the pattern for refining existing search results or identifying search results pertaining to "car reviews".

[0096] Pattern extraction is typically dynamic depending on the location identifier and the associated semantic type and/or attributes of the content of the object or of the various fields/segments in the pattern. In one embodiment, dynamic pattern extraction is automated using one or more algorithms. For example, each URL can be broken up into fragments (e.g., based on path segments and/or using the query parameters). The fragmented portions can be combined using various permutations. The permutations of the various combinations can then be filtered or selected based on various criteria. For example, the permutations with lesser number of segments that also match the URIs/URLs (location identifiers) of interest can be identified.

[0097] In some embodiments, the extraction module 306 extracts the pattern from the location identifier via user input. In some instances, the semantic types are user-specified or user defined and can be redefined or re-configured. Thus, the pattern for the user-specified semantic type can be automatically determined/extracted when the object (webpage) is identified by the user as having content of or related to the user-specified semantic type. Alternatively, the pattern for the user-specified semantic type may also be manually specified or modified by the user.

[0098] Note that multiple patterns can be determined for a single semantic type based on identification of multiple webpages as having content of or relating to the semantic type by multiple users. All or a portion of these patterns can be subsequently used for identifying

content of or relating to the semantic type. In some embodiments, the multiple patterns defined for a single semantic type can be ranked based on the number of users that have identified the associated content as having content of or relating to the semantic type. The rankings can be used in determining which patterns are subsequently used by default for search or search enhancement or for users who have not defined customized vocabulary of semantic types and/or attributes.

[0099] In one embodiment, the pattern corresponding to the semantic type also corresponds to an attribute of the semantic type of the content in the object (web page). For example, the pattern may correspond to the semantic type of 'car models' but also correspond to the attribute 'car reviews' of 'car models'. In general, patterns extracted by the extraction module 306 include a wildcard segment. The pattern may also include a domain name segment, a semantic type segment, and/or an attributes segment. An example of a location identifier pattern is illustrated with further reference to the examples of FIG. 11A-B.

[0100] In one embodiment, the extraction module 306 extracts a sub-pattern from the location identifier of the object (web page). The sub-pattern can correspond to attributes or other properties of the semantic type. For example, if the pattern "www.reviews.com/cars/\*" (where "\*" represents the wildcard segment of the pattern) corresponds to the semantic type of "cars", the sub-pattern of

"www.reviews.com/cars/reviews/\*" can be extracted by the extraction module 306 to correspond to the attribute of "car reviews" and stored for subsequent use, for example, by the search engine 314 in optimizing or performing search for content related to "car reviews". Multiple sub-patterns, each corresponding to one of multiple attributes of the semantic type can be extracted by the extraction module 306.

[0101] One embodiment of the host server 300 further includes a search engine 314. The search engine 314 can be any combination of software agents and/or hardware modules able to search, locate, identify, detect, objects having electronic content that satisfy one or more search criteria contained in a search query.

[0102] The search engine 314 can identify a set of search results in response to a search query via keyword search, query expansion, indexing content, meta-search, rankings, vertical search (e.g., domain-specific search), semantic search, or any combination of the

above, etc. In one embodiment, the search engine 314 performs search and/or search optimization using patterns identified from location identifiers (URLs or URIs) of objects (web pages) having content with defined semantic types. The pattern-facilitated search method and search optimization technique can be used in lieu of and/or in conjunction with any of the above listed search processes at any stage of the search process.

[0103] Note that the search engine 314 can identify search results from web pages or other types of electronic content from various sources. The sources can include objects distributed over the Internet (World Wide Web), objects stored locally, objects shared over a local or private network, etc. The sources can also include user content locally provided or shared by a network of users over networking site or multiple networking sites (e.g., a knowledge networking site, a semantics-enabled knowledge networking site, a social network site, a professional networking site, etc.). The user content that can be searched and analyzed by the search engine 314 and used to identify search results includes user-added content, user-generated content, and/or user supplemented content.

[0104] In one embodiment, the search engine 314 detects a set of location identifiers that have a pattern that matches a specified pattern. The specified pattern generally corresponds to a semantic type and can be stored in a computer-readable storage medium in the host server 300 or the knowledge repository 330 and retrieved for use. The search engine 314 can then identify a set of search results as having content related to the semantic type using the specified pattern. The set of search results can include the objects associated with the set of location identifiers having the specified pattern. The location identifier may be a Universal Resource Identifier (URI) or Universal Resource Locator (URL) and the objects may be web pages.

[0105] The search engine 314 can use, for example, a URI pattern to locate web pages with generally the same or similar pattern and use these web pages to generate a set of search results without using other search methods. The search engine 314 can also enhance the search relevancy by using URI patterns to categorize search results (e.g., can include search results identified from one or more other search methods such as keyword or indexed content search) as having content of or relating to a particular semantic type associated with a particular URI pattern, for example.

[0106] The search engine 314, using a stored pattern on a computer-readable storage medium in the host server 300, performs pattern recognition on multiple location identifiers and detects matching location identifiers having generally a same or similar pattern as the stored pattern. The pattern recognition may be performed by the recognition module 316 in the search engine 314 on location identifiers having same or different domain names from which the stored pattern was extracted.

[0107] Pattern recognition may initially be performed on location identifiers having the same domain names to identify a set of type-determined web pages having content of or related to the semantic type. An additional set of matching location identifiers can be detected using the additional location identifiers of a different domain name but having generally the same or similar pattern as the pattern extracted from the location identifier. In addition, the search engine 314 also performs pattern recognition to detect a subset of the matching location identifiers that have generally the same patterns as sub-patterns.

[0108] The search engine 314, using the matching location identifiers (from both the same or different domain names), can identify a set of type determined objects (e.g., type-determined web pages) having the matching location identifiers (e.g., URIs or URLs) and can generate a set of search results from the set of type-determined web pages. Note that each of the type-determined objects or web pages will typically have content of or related to the semantic type associated with the stored pattern used in identifying the matching location identifiers.

[0109] Furthermore, the search engine 314, using the subset of the matching location identifiers, further identifies a subset of the type-determined web pages having a subset of the matching URIs. Each of the subset of the type-determined objects or web pages includes content of or related to the attribute of the semantic type. In addition, multiple subsets of the type-determined web pages can be identified using multiple sub-patterns. Each of the multiple subsets of the type-determined web pages include content associated with one of the multiple attributes of the semantic type. The type-determined objects and/or web pages or the subsets thereof may be identified by the results generator module 318 in the search engine 314.

[0110] The search engine 314 can also use the matching location identifiers to refine an existing set of search results to identify objects having the content of or related to the

specific semantic type that corresponds to the stored pattern. For example, the search engine 314 can semantically categorize the existing set of search results and identifying a subset of which as having content related to the semantic type or one or more attributes of a semantic type in which the user is interested. In one embodiment, the semantic categorization is performed by the search optimizer module 320.

[0111] One embodiment of the host server 300 further includes a data mining module 322. The data mining module 322 can be any combination of software agents and/or hardware modules able to analyze and extract information from data through text mining, classification, topic detection, entity recognition, speech recognition, natural language processing or statistical natural language processing.

[0112] In one embodiment, the data mining module 322 data mines various objects (e.g., objects having electronic content) to determine what the content is or is related to. The objects can be web pages or other types of electronic content including but not limited to, digital music files, video files, electronic documents, multimedia content, etc. The results of the data mining can generally be used for search or in conjunction with other mechanisms of performing search.

[0113] In one embodiment, the data mining is used in conjunction with location identifier pattern extraction/matching for search and search optimization. For example, the data mining module 322 can data mine the content (e.g., interactive, textual, graphical, audio, etc.) of a set of type-determined web pages identified via pattern recognition using location identifier patterns. The data mining can be optionally performed to further determine relevancy to the semantic type to refine the search results identified by matching location identifier patterns, if needed.

[0114] Using the results of data mining, a refined set of type-determined web pages can be identified from the set of type-determined web pages based on the relevancy to the semantic type determined via the data mining. The refined set of type-determined web pages can be used to generate refined search results. In addition, using the results of data mining performed by the data mining module 322, each of the set of type-determined web pages can be ranked, based on the relevancy.

[0115] One embodiment of the host server 300 further includes a dialogue analyzer module 324. The dialogue analyzer module 324 can be any combination of software agents

and/or hardware modules able to parse, analyze, process, and extract information from a web-based dialogue.

[0116] The web-based dialogue analyzer module 324 can parse web-based dialogues according to a grammar (e.g., a grammar stored in the grammar module). In one embodiment, the grammar is specifically defined for parsing electronic conversations that occur, for example, in a social network via a web browser, online chat application, short-messages, etc. In general, web-based dialogues can include one or more of, online chat, a review, a comment, a status update, and/or a posting.

[0117] The electronic conversations or web-dialogues can be parsed, using the grammar, to extract information or knowledge regarding objects referred to in the conversation. For example, knowledge or information regarding video games, DVDs, books, CDs, etc. can be extracted from user reviews or comments on various web sites where these items are posted for discussion or for purchase.

[0118] In one embodiment, a web-based dialogue regarding an object (e.g., a webpage) is parsed, for example, using a grammar specifically defined for electronic or online conversations. Based on the parsing performed by the dialogue analyzer module 324, the object or web page can be identified as having content of or related to a semantic type. Based on the parsing, the object or web page may also be identified as having content or related to one or more attributes of the semantic type.

[0119] In one embodiment, the search engine 314 can further use the results of the dialogue parsing to semantically categorize the object or web page as having content of or related to an attribute or a semantic type. In one embodiment, the search optimizer module 320 of the search engine 314 performs the semantic categorization based on the results of the dialogue parsing.

[0120] One embodiment of the host server 300 includes a user module 304. The user module 304 can be any combination of software agents and/or hardware modules able to receive, manage, process, user search preferences and user-defined semantics (e.g., semantic types and attributes of semantic types) for search and search optimization.

[0121] For example, the user module 304 can track different semantic types and/or associated attributes defined by different users (e.g., a first set of semantic types defined by

a first user and a second set of semantic types defined by a second user, etc.). By tracking different semantic types/attributes defined by different users, user-customized search results can be optionally provided when content search performed for different users are based on their individually specified semantic types and attributes.

5 [0122] To generate user-customized search results, the search engine 314 can identify different sets of location identifier patterns (e.g., URI patterns) for different users. For example, the search engine 314 can identify a first set of URI patterns that are associated with the first set of semantic types defined by the first user and a second set of URI  
10 patterns that are associated with the second set of semantic types defined by the second user. Note that although the host server 300 is able to provide user-customized search results using user customized vocabulary, a default vocabulary can also be used to provide non-customized search when user-customization is not desired or not requested.

[0123] The first set of URI patterns can be identified from a first set of URIs of web pages having content specified by the first user as having content of the first set of semantic types  
15 and the second set of URI patterns can be identified from a second set of URIs of web pages having content specified by the second user as having content of the second set of semantic types.

[0124] For the first user, the search engine 314 can semantically categorize a first set of search results. The first set of search results can be categorized as having content related to  
20 one or more of the first set of semantic types that are defined by the first user. For the second user, the search engine 314 can semantically categorize a second set of search results. The second set of search results can be categorized as having content related to one or more of the second set of semantic types that are defined by the second user.

[0125] The first set of search results are categorized using the first set of URI patterns and  
25 the second set of search results are categorized using the second set of URI patterns. The search results can be categorized by the search optimizer module 320. In addition, the first set of search results are presented as being visually categorized to the first user as having content of or related to the one or more of the first set of semantic types and the second set of search results are presented as being visually categorized to the second user as having  
30 content related to the one or more of the second set of semantic types.



[0126] Note that the first and second sets of results may be categorized differently in response to the same keyword searches queried by the first and second users. For example, the first user may have defined the attributes "Bordeaux" and "1982 Vintage" of the semantic type "red wine". The second user may have defined "Burgundy" and "Tempranillo" of the semantic type "red wine". Therefore, when both users search for "red wine", the first user may see search results categorized under "Bordeaux" and "1982 Vintage" where the second user may see search results categorized under "Burgundy" and "1995 Vintage".

[0127] Alternatively, the attributes defined by various users are available (e.g., public) to all subsequent users. For example, if the first user defines attributes "Bordeaux" and "1982 Vintage" and the second user defines "Burgundy" and "Tempranillo", each of the four attribute types may be available to each other or another user that searches for "red wine". In some embodiments, users can select, from the available attributes, which attributes/properties they wish to search results for.

[0128] One embodiment of the host server 300 further includes a search engine builder 310. The search engine builder 310 can be any combination of software agents and/or hardware modules able to define, build, deploy, and/or maintain a search engine that is semantics enhanced.

[0129] The search engine builder 310 communicates with the location identifier retrieval module 304 to define a location identifier pattern for identifying objects having content of a semantic type. The location identifier pattern can be stored in a computer database embodied on a computer-readable storage medium. In one embodiment, the location identifier is a Universal Resource Identifier (URI) or Universal Resource Locator (URL) and the objects are web pages.

[0130] The search engine builder 310 can subsequently use the location identifier pattern to analyze a sitemap stored on a host server for a particular domain. To identify objects having content of or related to the semantic type, the search engine builder 310 identifies multiple location identifiers of the particular domain in the sitemap that have patterns that match the location identifier pattern defined for the semantic type.

[0131] Each of the multiple location identifiers that match the pattern can be indexed as being associated with objects having content of or related to the semantic type. In one

embodiment, each of the objects having location identifiers with matching patterns can be indexed as having content of or related to the semantic type. The indexes can optionally be used during searches to locate objects (e.g., web pages) having content of or related to various semantic types without having to perform pattern recognition on location

5 identifiers (e.g. URIs and/or URLs).

[0132] The search engine builder 310 can also analyze additional sitemaps or other domains to index the web pages. For example, the search engine builder 310 identifies, from the additional sitemaps, multiple additional location identifiers of the other domains matching the location identifier pattern defined for the semantic type. Each of the  
10 additional multiple location identifiers can be indexed as being associated with objects having content of or related to the semantic type.

[0133] In one embodiment, the search engine builder 310 defines a supplemental location identifier pattern for identifying objects having content related to an attribute of the semantic type. The search engine builder 310 can identify, from the sitemap, multiple  
15 supplemental location identifiers matching the supplemental location identifier pattern defined for the attribute. The sitemap can be the sitemap of the particular domain or other domains. Once the multiple supplemental location identifiers have been identified, the search engine builder 310 can index each of the multiple supplemental location identifiers as being associated with objects having content related to the attribute of the semantic type.

20 [0134] The indexes can be associated with semantic types and attributes and when indexed on objects or other content, can be used by the search engine 314 or other search engines for locating objects or web pages having content of or related to the attribute of the semantic type without performing further analysis.

[0135] One embodiment of the host server 300 further includes an advertisement module  
25 334. The advertisement module 334 can be any combination of software agents and/or hardware modules able to identify, refine, select, either semantically and/or based on keywords, advertisements using search results, item detail pages, and/or user feeds (e.g., user interest feeds).

[0136] The advertisement module 334 can use the semantic types and/or attributes  
30 identified in a set of search results to identify advertisement to provide enhanced ad targeting. In addition, advertisement can be targeted to search results or other types of

electronics objects according to semantic constraints including but not limited to semantic types/groups (e.g., drug, hotel, resume, event, laptop, vehicle, etc.) and/or semantic attributes/properties (e.g., cost, hours of operation, genre, vintage, year, horsepower, etc.).

[0137] For example, semantic types and/or attributes of search results can be used for filtering the associated advertisements. For example, when a user selects to view search results including content of or related to a particular semantic type (e.g., drug), the advertisements relevant to the semantic type "drug" can be presented to the user.

[0138] In addition to filtering advertisement based on semantic type, advertisements can further be filtered based on semantic attributes (e.g., semantic properties or facets) of the semantic type. For example, if the user selects to view search results related to semantic type "drug", the search results can further be filtered by its associated attribute (e.g., price, pediatric or adult, capsule, syrup, AM/PM, etc.). Based on the further refinement according to attributes, advertisements can also be refined. For example, when a user selects the pediatric attribute of the "drug" type, advertisements pertaining to children's medicine can be identified and presented to the user.

[0139] In one embodiment, when a user clicks on a selected search result, the user is re-directed to an item detail page (e.g., an ad) rather than to the URL of the selected search result. Each level of refinement (e.g., selection of semantic type, semantic attribute, and item detail page) of search results selected by the user allows the selection and identification of advertisements to also be further refined (e.g., semantically). The enhanced refinement increases the relevancy with what the user is searching for and thus increases the chances that the advertisement will be clicked on.

[0140] In general, the advertisements can include third party content, hosted content, and/or a combination of both. The advertisement module 334 can index both types of advertisement content (e.g., hosted or third party) according to semantic types or attributes for use locally or on third-party sites. The information related to the indexing of content and the relationship of the indices with the semantic types/contents can be provided to third-party hosts for a fee. In addition, the ad module 334 can index third-party content (ads) based on semantic content for a fee.

[0141] One embodiment of the host server 300 further includes a user interface module 312. The user interface module 312 can be any combination of software agents and/or

hardware modules able to generate one or more user interfaces for users to enter a search query to request search and to generate one or more user interfaces to display the search results.

5 [0142] In one embodiment, the user interface module 312 generates an interface (e.g., graphical user interface) for users to submit a search query and to access search (e.g., the example UI of FIG. 17) and/or optimized search. The search accessed via the interface (e.g., GUI of FIG. 17) is generally semantics-enhanced and may be user-customized. For example, the user can login before requesting search such that personalized semantic types and associated attributes (e.g., customized ontology) can be used in performing search  
10 and/or optimizing search results.

[0143] The search results can be displayed in a user interface and laid out according to the determined semantic type. Different search results having content of or related to different semantic types can be displayed on different pages accessible via a user interface. In one embodiment, the user interface module 312 presents a set of objects (e.g., web pages) of a  
15 semantic type on a first page and another set of objects of another semantic type on a second page, both accessible via the search results interface (e.g., examples of the various pages in the user interface of FIG. 18-23). In addition, the user may be able to access web pages having content of or related to attributes of the semantic type via the first page in the user interface.

20 [0144] For example, the user interface module 312 can present links to each of the multiple subsets of the type-determined web pages that are associated with attributes of the semantic type on the first page where the search results having content of or related the semantic type are displayed. Each of the links in the first page in the user interface may be identified by each of the multiple attributes of the semantic type.

25 [0145] In one embodiment, the user interface module 312 generates a template for differentiating search results from web pages hosted by sponsored sites (e.g., the search result 1952 in the example of FIG. 19B). Web pages of search results hosted by sponsored sites may also be prioritized by the user interface module 312, in the list of search results in the user interface. Sponsorships can be managed by the sponsorship module 326 which  
30 manages the partnerships with other hosts and any fees associated with the sponsorship.

[0146] In one embodiment, the user interface module 312 also provides user interfaces for the search engine builder 310 for managing location identifier patterns and their associated semantics types, for example. For example, through the example user interfaces shown in **FIG. 12-13**, location identifier patterns and their properties can be viewed and edited as needed. The semantic types associated with the patterns can also be viewed and updated through the user interface. The user interface provided for the search engine builder 310 can also provide statistical information including but not limited to, the number of URIs or URLs having a matching pattern.

[0147] The host server 300 represents any one or a portion of the functions described for the modules. More or less functions can be included, in whole or in part, without deviating from the novel art of the disclosure.

[0148] **FIG. 3B** depicts an example block diagram illustrating the components of the host server 300 that performs search and/or optimized search.

[0149] In one embodiment, host server 300 includes a network interface 302, a processing unit 334, a memory unit 336, and/or a storage unit 338. Additional or less units or modules may be included. One suitable network interface 302 has been described in the example of **FIG. 3A**.

[0150] One embodiment the host server 300 further includes a processing unit 334. The data received from the network interface 302 can be input to the processing unit 334. The data that is received can include search queries, content from various content sources or a user content repository. The processing unit 334 can include one or more processors, CPUs, microcontrollers, FPGAs, ASICs, DSPs, or any combination of the above. Data that is input to the host server 300 can be processed by the processing unit 334 and output to a display and/or output via a wired or wireless connection to an external device, such as a mobile phone, a portable device, a host or server computer by way of a communications component.

[0151] One embodiment of the host server 300 further includes a memory unit 336 and a storage unit 338. The memory unit 335 and a storage unit 338 are, in some embodiments, coupled to the processing unit 334. The memory unit can include volatile and/or non-volatile memory. In performing search and search optimization (e.g., semantics-enhanced search optimization), the processing unit 334 may perform one or more processes related

to pattern extraction in a location identifier of a source of content and processes related to pattern recognition in identifying objects (e.g., web pages or other types of electronic source) having content of or related to specific semantic types or attributes based on the specific search query.

5 [0152] In some embodiments, any portion of or all of the functions described of the various example modules in the host server 300 of the example of **FIG. 3A** can be performed by the processing unit 334. In particular, with reference to the host server illustrated in **FIG. 3A**, the functions and techniques executed by the location identifier retrieval module, the extraction module, the user module, the search engine, the  
10 recognition module, the results generator module, the search optimizer module, the data mining module, the dialogue analyzer module, the user interface module, the advertisement module, and/or the sponsorship module can be performed via any of the combinations of modules in the control subsystem that are not illustrated, including, but not limited to, the processing unit 334 and/or the memory unit 336.

15 [0153] **FIG. 4** depicts a flow chart illustrating an example process for optimizing search by identifying web pages having content of or related to a semantic type.

[0154] In process 402, a location identifier of a webpage having content associated with a semantic type is identified. The location identifier may be a Universal Resource Identifier (URI) or Universal Resource Locator (URL). In process 404, a pattern is extracted from  
20 the location identifier of the webpage and can be stored in a computer database embodied in a computer-readable storage medium. One suitable process for optimizing search by identifying web pages having content associated with one of multiple attributes of the semantic type is described in more detail in the example illustrated in **FIG. 5**.

[0155] In process 406, pattern recognition is performed on multiple location identifiers  
25 having a same domain name as the location identifier using the pattern stored on the computer-readable storage medium. The multiple location identifiers may be identified from a sitemap. In process 408, matching location identifiers from the multiple location identifiers are detected as having a same pattern as the pattern extracted from the location identifier.

30 [0156] In process 410, a set of type-determined web pages having the matching location identifiers are identified. In one embodiment, each of the set of type-determined web-pages

is associated with an ontology defining the semantic type. In process 412, the set of type-determined web pages are presented on a first page in a user interface as categorized as having the semantic type. Note that another set of type-determined web pages can be presented on a second page of the user interface is categorized as having another semantic type.

[0157] In process 414, a set of search results are generated from the set of type-determined web pages. In one embodiment, additional semantic data is extracted from HTML or XHTML content of each of the set of type-determined web pages (e.g., using GRDDL). The additional semantic data can be used generate a refined set of search results from the set of search results using the additional semantic data.

[0158] In process 416, textual content of each of the set of type-determined web pages is data mined to further determine relevancy to the semantic type. In process 418, each of the set of type-determined web pages is ranked based on the relevancy determined from the data mining. In process 420, a refined set of type-determined web pages is generated from the set of type-determined web pages based on the relevancy to the semantic type determined via the data mining. In process 422, a refined set of search results is generated from the refined set of type-determined web pages.

[0159] FIG. 5 depicts a flow chart illustrating an example process for optimizing search by identifying web pages having content associated with multiple attributes of the semantic type.

[0160] In process 502, a sub-pattern is extracted from the location identifier of the webpage. In one embodiment, the sub-pattern corresponds to an attribute of the semantic type of the content. In process 504, pattern recognition is performed on the multiple location identifiers using the sub-pattern. In process 506, a subset of the matching location identifiers having a same pattern as the sub-pattern are detected. In process 508, a subset of the type-determined web pages having the subset of the matching URIs is identified. In general, each of the subset of the type-determined web pages include content associated with the attribute of the semantic type.

[0161] In process 510, multiple sub-patterns are extracted from the location identifier. In process 512, multiple subsets of the set of search results including multiple subsets of the type-determined web pages are identified using the multiple sub-patterns. In process 514,

links to each of the multiple subsets of the type-determined web pages that are associated with one of the multiple attributes of the semantic type are presented on the first page in the user interface.

**[0162] FIG. 6** depicts a flow chart illustrating an example process for search of objects having content of or related to a semantic type using patterns of the location identifiers of the objects.

**[0163]** In process 602, a set of location identifiers that have a pattern that matches a specified pattern are identified. In general, the specified pattern corresponds to a semantic type. In process 604, set of search results as having content related to the semantic type are identified. The set of search results can include objects associated with the set of location identifiers having the specified pattern. The objects can be linked to an ontology defining the semantic type and/or attributes (e.g., properties) of the semantic type. In one embodiment, the location identifier is a Universal Resource Identifier (URI) or Universal Resource Locator (URL) and the objects are web pages.

**[0164]** In process 606, textual content of each of the set of search results is data mined to further determine relevancy to the semantic type. In process 608, a refined set of search results are identified from the set of type-determined web pages based on the relevancy to the semantic type determine via the data mining. In process 610, each of the set of search results are ranked based on the relevancy determined from the data mining.

**[0165] FIG. 7** depicts a flow chart illustrating an example process for optimizing search using semantics.

**[0166]** In process 702, a Universal Resource Identifier (URI) of a webpage having content associated with a semantic type as specified by a user is identified. The semantic type can be defined by the user through tagging the webpage on a user interface. In process 704, a pattern that corresponds to the semantic type of the content in the webpage is identified from the URI of the webpage, based on the semantic type that is specified by the user.

**[0167]** In process 706, additional patterns for the semantic type are identified based on identification of additional webpage(s) as having content of the semantic type by the user. In one embodiment, each of the additional patterns identified for the semantic type are



ranked based on a number of users that identified each of the additional web pages as having content of the semantic type.

[0168] In some instances, users can belong to different levels (e.g., topic experts, power users, normal users, guests, etc.). The level differentiation can provide users with different access levels to tagging web pages or other objects with semantic types. The level differentiation can also be used by the system to determine the user-definitions to adopt or modify as the default semantic type for objects. Further, the level differentiation can be used by the system to determine which user-defined/specified tags for objects or web pages are provided to other users for reference and/or to adopt as their own. In general, users who have been specified as topic experts or ranked/reviewed by other users as providing quality definitions are assigned a higher level. For example, their definitions of objects may be provided to other users and may be considered with heavier weight by the system when aggregating various user definitions.

[0169] In process 708, the pattern and the additional patterns are used to perform pattern recognition on multiple URIs. In process 710, matching URIs having a same pattern as the pattern extracted from the URI are detected from the multiple URIs.

[0170] In process 712, a set of type-determined web pages having the matching URIs are identified. In process 714, each of the additional patterns identified for the semantic type is ranked based on a number of users that identified each of the additional web pages as having content of the semantic type.

[0171] In one embodiment, the webpage is identified by a second user as having content of a second semantic type that is different from the semantic type. A second pattern that corresponds to the second semantic type of the content in the webpage, can be identified from the URI of the webpage based on the second semantic type that is specified by the second user. The second pattern may be different from or same as the pattern. The second pattern can thus be used for identifying a second set of type-determined web pages having content that is of or related to the second semantic type.

[0172] FIG. 8 depicts a flow chart illustrating an example process for generating user-customized search results using user-defined semantic types.

[0173] In process 802, a first set of URI patterns that are associated with a first set of semantic types defined by a first user are identified. In one embodiment, the first set of URI patterns are identified from a first set of URIs of web pages having content specified by the first user as having content of the first set of semantic types.

5 [0174] In process 804, using the first set of URI patterns, a first set of search results are semantically categorized for the first user, as having content related to one or more of the first set of semantic types defined by the first user. The first set of search results can be categorized by comparing associated URIs of each of the first set of search results with the first set of URI patterns.

10 [0175] In addition, a first additional set of URI patterns that are associated with a first set of attributes of a semantic type that are defined by the first user are identified. Using the first additional set of URI patterns, the first set of search results can be categorized for the first user, as having content related to one or more of the first set of attributes of the semantic type. In general, the first additional set of URI patterns are identified from a first  
15 additional set of URIs of web pages having content specified by the first user as having content related to one or more of the first set of attributes of the semantic type.

[0176] In process 806, a web-dialogue regarding a particular webpage in which the first user is participating is parsed according to a grammar. In process 808, the particular web page, is categorized, based on the parsing, as having content related to one or more of the  
20 first set of semantic types defined by the first user.

[0177] In process 810, a second set of URI patterns that are associated with a second set of semantic types defined by a second user are identified. In one embodiment, second set of URI patterns are identified from a second set of URIs of web pages having content specified by the second user as having content of the second set of semantic types. In  
25 process 812, using the second set of URI patterns, a second set of search results are semantically categorized for the second user, as having content related to one or more of the second set of semantic types defined by the second user.

[0178] In one embodiment, the first set of search results are identified in response to a keyword search requested by the first user and the second set of search results are  
30 identified in response to the keyword search requested by the second user that is a same keyword search as the first user's keyword search. The first set of search results that are

presented can be visually categorized to the first user as having content related to the one or more of the first set of semantic types. The second set of search results that are presented can be visually categorized to the second user as having content related to the one or more of the second set of semantic types

- 5   **[0179] FIG. 9** depicts a flow chart illustrating an example process for building a semantics-enhanced search engine.

10   **[0180]** In process 902, a location identifier pattern for identifying objects having content of a semantic type is defined. The location identifier pattern can be stored in a computer database embodied on a computer-readable storage medium. In process 904, a sitemap stored on a host server for a particular domain is analyzed. In process 906, multiple location identifiers of the particular domain matching the location identifier pattern defined for the semantic type are identified from the sitemap. In process 908, each of the multiple location identifiers is indexed as being associated with objects having content of the semantic type.

15   **[0181]** In one embodiment, the location identifier is a Universal Resource Identifier (URI) or Universal Resource Locator (URL) and the objects are web pages. The web pages can be associated with an ontology defining the semantic type. In addition, the URL/URI can be associated with the ontology. In one embodiment, the HTML or XHTML content in the web pages are converted to semantic data for use in the semantics-enhanced search engine.  
20   The HTML or XHTML content can be converted using the GRDDL (Gleaning Resource Descriptions from Dialects of Languages) markup format.

25   **[0182]** In process 910, a grammar for parsing web-based dialogue is defined. The grammar can be a general grammar or a domain-specific grammar. In process 912, a web-based dialogue regarding a particular object is parsed according to the grammar. In process 914, the particular object is identified as having content of the semantic type based on the parsing. In one embodiment, a predicate and object of the web-based dialogue are detected using the grammar. RDF triplets may also be generated from the web-based dialogue and can be used with the grammar for conversational search in the semantics-enhanced search engine. In general, a web-based dialogue can include one or more of, a review, a comment,  
30   a status update, and/or a posting.

[0183] In process 916, additional sitemaps of other domains are analyzed. In process 918, multiple additional location identifiers of the other domains matching the location identifier pattern defined for the semantic type are identified from the additional sitemaps. In process 920, each of the additional multiple location identifiers are indexed as objects  
5 having content of the semantic type.

[0184] Furthermore, a supplemental location identifier pattern for identifying objects having content related to an attribute of the semantic type is identified. From the sitemap, multiple supplemental location identifiers of the particular domain matching the supplemental location identifier pattern defined for the attribute of the semantic type can  
10 be identified. Each of the multiple supplemental location identifiers can then be indexed as being associated with objects having content related to the attribute of the semantic type.

[0185] **FIG. 10** depicts a flow chart illustrating an example process for performing a search by mapping a search query into a specified pattern in a pattern template.

[0186] In process 1002, a pattern template followed in one or more URIs is identified. The  
15 pattern template can include multiple specified patterns which can be used to resolve a plurality of semantic object types. In process 1004, the pattern template is stored in a computer database embodied in a computer-readable storage medium for use in subsequent searches. In process 1006, a user interface that enables a user to perform searching is provided. IN process 1008, a search query is received via the user interface. In process  
20 1010, the pattern template is retrieved from the computer database.

[0187] In process 1012, the search query is mapped into a specified pattern of the multiple specified patterns. In process 1014, a search for matching URIs having a pattern that matches the specified pattern is performed. Web pages corresponding to the matching URIs can be presented as search results to the user.

25 [0188] **FIG. 11A-B** illustrate examples of location identifier patterns.

[0189] In the example of **FIG. 11A**, an example of the location identifier pattern 1100 includes a domain name segment 1102, a semantic type segment 1106, and a wildcard segment 1104. In some embodiments, location identifier patterns can also include an attribute type segment (not shown). The location identifier pattern 1100 can be used for  
30 identifying web pages from wikitravel.org having a URL beginning with

"wikitravel.org/africa". The wildcard segment 1104 appending the segment 1102 and "/africa" indicates that any webpage having a URL or URI with a beginning matching "wikitravel.org/africa" can be identified as matching the pattern 1100.

[0190] In the example of **FIG. 11B**, the location identifier pattern 1100 does not include the domain name 1112. The location identifier pattern 1110 includes a semantic type segment 1116 and a wildcard segment 1104. The location identifier pattern 1110 can be used for identifying web pages not limited to wikitravel.org but having the word 'africa' in the URI or URL.

[0191] **FIG. 12A** illustrates an example screenshot 1200 showing examples of location identifier patterns and associated semantic types.

[0192] The example location identifiers patterns 1204 and 1206 illustrated are extracted from web pages in the domain of wikitravel.org. Each of the location identifier patterns 1204 and 1206 can be associated with a semantic type. For example, the semantic type 1208 "Travel" is associated with the pattern 1206 and the semantic type 1210 "History" is associated with the pattern 1204. The pattern 1204 can thus be used to identify objects or web pages with patterns matching 1204. These identified web pages can be used in search results in response to a query for items of the semantic type "History". Similarly, search results related to the semantic type "Travel" can be identified using the pattern 1206.

[0193] The screenshot also shows statistics related to the identified or specified patterns (e.g., pattern 1204 and pattern 1206). For example, the screenshot shows the number of sitemap URLs that match the patterns 1204 and 1206 and when the sitemap was last scanned. Note that although in one embodiment, sitemap scans are used to identify matching URIs or URLs, other methods can be used to obtain a list of URIs and URLs for hosts (e.g., site crawling).

[0194] In other embodiments, URIs and URLs can be manually identified or seeded from other resources. For example, a web crawler (e.g., a web crawl corpus such as Build Your Own Search Service (BOSS) platform) can be used to filter domains. In addition, a manual crawl of the site can be performed to extract URLs and/or URIs in domains. Other known and/or convenient methods can also be used to extract location identifiers.

[0195] **FIG. 12B** illustrates an example screenshot 1225 of a location identifier 1204 and various editable fields.

[0196] The editable fields allow various segments or portions of the pattern to be defined. The field "Group #" field 1216 identifies the specified portion of the pattern and the  
5 "Property" field 1218 indicates the semantic property of the value in the portion specified by the "Group #". For example, the Group #1 "en" in the pattern 1204 can be defined as having a semantic property of "language" or "English language". The group mappings 1214 are dynamic and generally variable among URIs and URLs.

[0197] The "Transform" field 1220 can be used to transform a value in a portion of the  
10 pattern to another format (which may be uniform for a particular type of data). For example, a date may be represented as 20090406 or 04062009 or 040609, etc. So a transform of "date ('yyyyMMdd')" would allow the system to properly parse it. In one embodiment, the value is transformed using Resource Description Framework (RDF) data model or other metadata data models suitable for modeling information in web resources.  
15 The transformation can be performed using GRDDL (Gleaning Resource Descriptions from Dialects of Languages) to enable RDF triples to be extracted from XML or (X)HTML documents.

[0198] The transformed value which can be a uniform or predetermined format allows for indexing based on semantic facets for search enhancement for example, by indexing the  
20 semantic content. The transformed value is generally machine-readable to enable semantic-indexing according to semantic facets which are generally user readable. Therefore, the transformation can also enable human users to make queries according to these semantic facets.

[0199] The static mapping field 1222 allows values or types to be assigned to various  
25 properties of location identifiers or objects associated with the location identifiers matching a particular pattern (e.g., pattern 1204). For example, a "Property" field 1224 can be assigned "topic" and the "Value" field 1226 can be assigned "Travel" (e.g., topic = travel for each web page having a URL with a matching pattern). Note that multiple properties can be defined for any given web page having the matching pattern.

30 [0200] The "knowledge transform" field 1228 allows an XSL document to be created. The XSL document can be used to detect values (e.g., values having semantic property) from

the HTML of any URL with a matching pattern. In this example, the transform would be applied against the matching travel review articles at wikitravel.org. The transform can identify the values (e.g., values having semantic properties) from the HTML source file of the web page that are not in the pattern of its associated URL or URI. In this manner,  
5 additional semantic information about the web page identified using pattern recognition can be extracted.

[0201] FIG. 12C illustrates another example screenshot 1250 of another location identifier 1206 and various editable fields. The various editable fields can similarly be used for pattern group mappings and static mappings.

10 [0202] FIG. 13A illustrates another example screenshot 1300 showing examples of location identifier patterns 1304 and 1306 and an associated semantic type 1308 for the domain 1302.

[0203] The screenshot 1300 is illustrated as an example for the domain 1302. The partnership level for the host of the domain 1302 can also be specified. In general, three  
15 partnership levels are available to various hosts (e.g., none, sponsored, or hosted).

[0204] Hosted content can be used to indicate that content is locally hosted as opposed to remotely indexed content. For example, the hosted content can be locally hosted by the search engine (e.g., a knowledge networking site hosted by the search engine). Hosted content can be assigned higher rankings or priorities when as a search result in a list of  
20 search results.

[0205] When the partnership level of a host site is "sponsored", the web pages from the host site can be displayed with priority in a user interface containing search results. In addition, web pages from a sponsored host site can be displayed with additional graphics or differentiating template (e.g., search result 1952 in the example of FIG. 19B). For  
25 example, the template for sponsoring sites can be defined under the web resources portion 1314. The template can include additional graphics, differentiating font, interactive features, audio , etc.

[0206] An example of a template for specifying how a search result having content from a specific domain of a particular semantic type can be displayed in a user interface is  
30 depicted FIG. 14A. The template is generally the HTML markup for a sponsored site. An

example of a stylesheet for specifying how a search result having content from a specific domain of a particular semantic type is depicted in **FIG. 14B**. The stylesheet can include the corresponding CSS markup for stylizing the sponsored site.

5 [0207] In this example, the associated semantic type 1308 "Disease" is specified for both patterns 1304 and 1306. In other words, web pages located using patterns 1304 and 1306 can be identified as having content of or related to the semantic type 1308 "Disease". Note that additional semantic types can be defined by selecting the "New" tab 1310.

10 [0208] **FIG. 13B** illustrates an example screenshot 1325 of location identifier 1304 and various editable fields. **FIG. 13C** illustrates another example screenshot 1350 of a location identifier 1306 and various editable fields.

[0209] **FIG. 15** illustrates an example list of domains 1502 for which location identifier patterns have been extracted.

15 [0210] The list of domains 1502 can be revised or added to. Each of the listed domains 1502 can have patterns defined for various semantic types. The defined patterns can be stored for each domain and the corresponding sitemaps can be scanned automatically or upon manual request to identify, from each domain, the URLs or URIs that match the defined patterns. Note that the defined patterns can be specific to a particular domain or can be used across different domains.

20 [0211] The list of domain names 1502 may also be associated with a list 1504 identifying the partnership level of each domain in the list 1502. For example, certain domains may be sponsors and identified as such under list 1504.

[0212] **FIG. 16A** illustrates an example list of vocabulary 1602 showing semantic types 1604 that have been defined. **FIG. 16B** illustrates an example list of vocabulary properties 1606 showing a list of attributes 1608 of the semantic types that have been defined.

25 [0213] **FIG. 17** illustrates an example screenshot of a user interface of a search engine that executes search and/or optimizes search by performing pattern recognition using location identifier patterns.

[0214] The user interface includes a search field 1702 where search queries can be entered by a user. The search query may be used to initiate a keyword-based search, a semantics



search, or a combination thereof. The search can trigger identification of web pages or other objects using the disclosed pattern recognition method along or in conjunction with other search mechanisms.

[0215] In one embodiment, user-customized search can be accessed by logging in as an identified user 1701. Once logged in, any user-defined semantic types and/or semantic type attributes can be used by the search engine in performing search or search optimization. For example, search results can be categorized on a user-by-user basis according to their defined semantic types and/or semantic type attributes.

[0216] FIG. 18 illustrates a screenshot 1800 showing an example list of search results identified using keyword search without further semantic-categorization.

[0217] The example list of search results is shown when the 'all' tab 1801 is selected. This example list includes results identified in response to a search query 'fever'. The search result entry 1850 is identified, based on the pattern of the associated identifier, as having content related to the semantic type "disease". The search result entry 1850 can be labeled as corresponding to the semantic type "Disease" 1810 in the user interface showing the search results.

[0218] In the user interface displaying search results, each or some of the search result entries can be associated with various tabs (e.g., "Review this" 1802, "Share this" 1804, "Discuss this" 1806, "Extract pattern" 1808). The user can review the result using tab 1802. Tab 1804 can be used by the user to share the entry with other users or to post the associated web page or other object on another website, for example. Tab 1806 can be used to initiate or participate in a discussion regarding the entry 1850.

[0219] Tab 1808 "Extract pattern" can be used by the user to extract the pattern of the associated location identifier (e.g., the URI or URL of a web page). Upon selection of tab 1808, the pattern can be automatically extracted by the system and stored in memory. The pattern, after extraction by the system, can also be edited by the user. In addition, the user can define the semantic type associated with the pattern. In one embodiment, the tab "What's this" can be selected by the user to specify the semantic type that is associated with content of the search result.

[0220] FIG. 19A illustrates a screenshot 1900 showing an example list of search results identified responsive to a search query and categorized as having content related to a semantic type 1901.

[0221] When tab 1901 "Diseases" is selected, the search results shown in the example of FIG. 18 having content or related to the semantic type "Diseases" are shown and categorized as such. The entry 1950 can be labeled in the user interface as having a semantic type of "Disease" 1910. In addition, a user can review the entry using tab 1902, share the entry using tab 1904, discuss the entry using tab 1906, and/or extract the pattern from the location identifier of the web page using tab 1908.

[0222] In one embodiment, when the semantic type "Diseases" are selected by the user, advertisements relevant to the selected semantic type can also be identified. For example, Ad 1 1922 for a book/magazine titled "Healthcare for Babies" can be presented to the user. In addition, Ad 2 1924 for an encyclopedia entitled "Encyclopedia of Fevers" can be identified and selected for presentation to the user.

[0223] FIG. 19B illustrates a screenshot 1925 showing an example of a search result 1952 hosted by a sponsored domain and identified as having content related to a semantic type.

[0224] The search result 1952 hosted by the sponsored domain can be displayed in the user interface as having an enhanced or otherwise differentiating template (e.g., enlarged font size, additional graphics or colors, etc.). In addition, search results from sponsored domains may be placed before other search results in the list of search results.

[0225] FIG. 20 illustrates a screenshot 2000 showing an example list of search results identified responsive to the search query and categorized as a news article.

[0226] When tab 2001 "News" is selected, the search results shown in the example of FIG. 18 having content of or related to a news article are shown and categorized as such. Note that the semantic type of entry 2050 may not be determined yet. The user can use the "What's this?" tab 2002 to enter a user-defined semantic type of the content of the search result 2050. The user-defined semantic type can be saved by the search engine and subsequently used to identify other content of or relating to the user-defined semantic type.

[0227] In one embodiment, the user can use the "extract pattern" tab 2008 to define a pattern for the user-defined semantic type. The defined pattern can also be saved and

subsequently used to identify additional objects with content of or related to the user-defined semantic type.

[0228] Note that the system can aggregate the various semantic types specified by users for the same search result 2050. In one embodiment, the system uses the user definitions to assign a semantic type to the search result. For example, the various semantic types defined by user can be ranked based on the frequency of use and the system can adopt one or more of the higher ranking semantic types to be assigned to the web page.

[0229] FIG. 21 illustrates a screenshot 2100 showing an example list of search results identified responsive to the search query and categorized as having content related to another semantic type 2101.

[0230] When tab 2101 "Drugs" is selected, the search results shown in the example of FIG. 18 having content or related to the semantic type "Drugs" are shown and categorized as such. The entry 2150 can be labeled in the user interface as having a semantic type of "Drugs" 2101. In addition, a user can review the entry using tab 2102, share the entry, discuss the entry, and/or extract the pattern from the location identifier of the web page using tab 2108.

[0231] In one embodiment, when the semantic type "Drugs" is selected by the user, advertisements relevant to the selected semantic type can also be identified. For example, Ad 1 2122 for Tylenol can be presented to the user. In addition, Ad 2 2124 for an encyclopedia entitled "Encyclopedia of Over-the-counter Drugs" can be identified and selected for presentation to the user.

[0232] FIG. 22 illustrates a screenshot 2200 showing an example list of search results identified responsive to the search query and categorized as having content related to yet another semantic type 2201.

[0233] When tab 2201 "Medical Conditions" is selected, the search results shown in the example of FIG. 18 having content or related to the semantic type "Medical Conditions" are shown and categorized as such. The entry 2150 can be labeled in the user interface as having a semantic type of "Medical Conditions" 2201. In addition, a user can review the entry using tab 2202, share the entry using 2204, discuss the entry using 2206, and/or extract the pattern from the location identifier of the web page using tab 2208.

[0234] FIG. 23 illustrates a screenshot 2300 showing an example list of search results identified responsive to the search query and categorized as having content related to yet another semantic type 2301.

5 [0235] When tab 2301 "Products" is selected, the search results shown in the example of FIG. 18 having content or related to the semantic type "Products" are shown and categorized as such. The entry 2350 can be labeled in the user interface as having a semantic type of "Products" 2301 and labeled as having the semantic type of "Product" 2310. In addition, a user can review the entry using tab 2302, share the entry using 2304, discuss the entry using 2306, and/or extract the pattern from the location identifier of the  
10 web page using tab 2308.

[0236] In one embodiment, when the semantic type "Products" is selected by the user, advertisements relevant to the selected semantic type can also be identified. For example, Ad 1 2322 for Tylenol can be presented to the user. In addition, Ad 2 2324 for Robitussin can be identified and selected for presentation to the user. Further, Ad 3 2326 for Aspirin  
15 can also be identified and selected for presentation to the user.

[0237] FIG. 24 illustrates an example of a web-dialogue 2404 that is parsed to obtain information regarding a particular webpage 2402.

[0238] The webpage 2402 includes a rating and a review of a movie. The web-dialogue 2404 conducted between users John and Mary are about the movie, other recently released  
20 movies, and recommendations, etc. The web-dialogue 2404 can be parsed according to a grammar. The grammar may be specifically designed for parsing online conversations or electronics conversations. By parsing the dialogue 2404, the semantic type of the content of webpage 2402 can be determined and used by the search engine in identifying search results or optimizing a set of search results.

25 [0239] In addition, triplets for conversations can be generated for a fixed subject and/or a variable subject. The fixed subject can be the content or resource being commented upon in the conversation (e.g., in reference to a webpage referencing "Pulp Fiction", the comment can be "this movie was directed by Quentin Tarantino"). The predicate and objects in the conversation can be detected using grammars which can be either a general  
30 grammar of a domain-specific grammar (e.g., movies or entertainment related).

[0240] The semantic facets identified using the predicate/objects can be used for targeted domain and/or conversational search (e.g., clicking on "movies", then clicking "Quentin Tarantino" under a "Director" filter; or typing in "who directed pulp fiction?"). In addition, the grammar (e.g., general or domain-specific) can also be used for targeted search or conversational search.

[0241] FIG. 25 depicts a flow chart illustrating an example process for semantic search enhanced advertising.

[0242] In process 2502, a search result is identified in response to a search query submitted by a user through a user device (e.g., computer, portable device, PDA, iPhone, cell phone, etc.). The search results can be identified using any method including but not limited to, a keyword search, indexing, semantics search, and/or any combination of the above. The search result can include a web page or other types of electronic content. In general, the search result is presented to a user interface for access by the user on the user device.

[0243] In process 2504, a semantic-type category with which content of a search result relates is identified. The semantic-type category can be identified via any method including but not limited to, topic detection, Natural Language Processing, data mining, speech analysis, speech recognition, or any combination of the above. In one embodiment, the semantic-type category is identified based on a pattern of a location identifier (e.g., a URI and/or a URL) of the web page.

[0244] In process 2506, one or more advertisements are selected in accordance with the semantic-type category with which the content of the search result relates. The advertisements may be selected by keyword matching using the semantic-type category. In process 2508, the one or more advertisements are presented to the user in the user interface.

[0245] Through the user interface, a selection of the search result by the user can be received. In one embodiment, instead of directing the user to a target page URL of the search result, a web page including item details is retrieved based on the semantic-type category for display to the user. The one or more advertisements can be presented to the user in the web page.

[0246] In one embodiment, a set of search results are identified responsive to the search query. Each of set of search results can be, for example, identified as having content related to a specific semantic-type category of a set of semantic-type categories.

Advertisements to be associated with each of the set of search results can be identified  
5 using the specific semantic-type category with which the content is related.

[0247] When a user request to view a specific search result of the search results is received, an advertisement can selected using the specific semantic-type category with which content of the specific search result relates. The selected advertisement can be presented to the user in conjunction with the specific search result.

10 [0248] In addition, each of the advertisements can be categorized according to the specific semantic-type category used to identify each of the advertisements. In one embodiment, an index is assigned to each of the advertisements according to the specific semantic-type. The indices can subsequently be stored for each of the advertisements in a machine-readable storage medium for subsequent usage, for example.

15 [0249] The advertisements can generally include third-party content, hosted content, or a combination of the above. In one embodiment, a third party host of the third-party content is charged for indexing each of the advertisements based on the specific semantic-type. Further, third-party hosts can be provided with the advertisements and the indices associated with each of the advertisements. In some instances, the third-party host is  
20 charged for the indices associated with each of the advertisements.

[0250] In one embodiment, at least a portion of the set of search results is identified as having content related to a specific attribute of the specific semantic-type category. The specific attribute can be determined using patterns of location identifiers of each of the set of search results. Advertisements to be associated with each of the at least a portion of the  
25 set of search results can further be identified using the specific attribute with which the content is related.

[0251] Each of the advertisements can be categorized according to the specific attribute of the specific semantic-type category used to identify each of the advertisements. In addition, an index can be assigned to each of the advertisements according to the specific attribute of the semantic-type category. The indices for each of the advertisements can be  
30 stored in a machine-readable storage medium.

**[0252] FIG. 26** depicts a flow chart illustrating an example process for targeting advertisements using semantic constraints.

**[0253]** In process 2602, a search result is identified in response to a search query submitted by a user. In process 2604, a semantic group with which content of a search result relates is identified. In process 2606, a facet of the semantic group with which the content of the search result relates is further determined. In one embodiment, the semantic group (e.g., also referred to as 'semantic type') is identified using a pattern of a location identifier of the search result. The semantic facet (e.g., also referred to as an attribute of the semantic type or group) can also be identified using the pattern of the location identifier of the search result. The search result generally includes a web page. The location identifier can be a URI and/or a URL.

**[0254]** In process 2608, an advertisement is identified based on the facet of the semantic group.

**[0255]** In process 2610, an index is assigned to the advertisement as being associated with the facet of the semantic group. In one embodiment, the index of the advertisement is provided to a third-party for a fee. In addition, the advertisement is hosted by a third party and the index can be assigned to the advertisement for the third-party for a fee.

**[0256]** In process 2612, the index of the advertisement is stored in a machine-readable storage medium. The index of the advertisement can be retrieved from the machine-readable storage medium to determine the facet with which the advertisement is associated.

**[0257]** In process 2614, the advertisement is presented to the user for viewing in a user interface in conjunction with the search result. In addition, in response to determining that a second search result includes content related to the facet of the semantic group, the same advertisement can be presented with the second search result.

**[0258] FIG. 27** shows a diagrammatic representation of a machine in the example form of a computer system 2700 within which a set of instructions, for causing the machine to perform any one or more of the methodologies discussed herein, may be executed.

**[0259]** In alternative embodiments, the machine operates as a standalone device or may be connected (e.g., networked) to other machines. In a networked deployment, the machine

may operate in the capacity of a server or a client machine in a client-server network environment, or as a peer machine in a peer-to-peer (or distributed) network environment.

[0260] The machine may be a server computer, a client computer, a personal computer (PC), a tablet PC, a laptop computer, a set-top box (STB), a personal digital assistant (PDA), a cellular telephone, an iPhone, a Blackberry, a processor, a telephone, a web appliance, a network router, switch or bridge, or any machine capable of executing a set of instructions (sequential or otherwise) that specify actions to be taken by that machine.

[0261] While the machine-readable medium or machine-readable storage medium is shown in an exemplary embodiment to be a single medium, the term "machine-readable medium" and "machine-readable storage medium" should be taken to include a single medium or multiple media (e.g., a centralized or distributed database, and/or associated caches and servers) that store the one or more sets of instructions. The term "machine-readable medium" and "machine-readable storage medium" shall also be taken to include any medium that is capable of storing, encoding or carrying a set of instructions for execution by the machine and that cause the machine to perform any one or more of the methodologies of the presently disclosed technique and innovation.

[0262] In general, the routines executed to implement the embodiments of the disclosure, may be implemented as part of an operating system or a specific application, component, program, object, module or sequence of instructions referred to as "computer programs."

The computer programs typically comprise one or more instructions set at various times in various memory and storage devices in a computer, and that, when read and executed by one or more processing units or processors in a computer, cause the computer to perform operations to execute elements involving the various aspects of the disclosure.

[0263] Moreover, while embodiments have been described in the context of fully functioning computers and computer systems, those skilled in the art will appreciate that the various embodiments are capable of being distributed as a program product in a variety of forms, and that the disclosure applies equally regardless of the particular type of machine or computer-readable media used to actually effect the distribution.

[0264] Further examples of machine-readable storage media, machine-readable media, or computer-readable (storage) media include but are not limited to recordable type media such as volatile and non-volatile memory devices, floppy and other removable disks, hard



disk drives, optical disks (e.g., Compact Disk Read-Only Memory (CD ROMS), Digital Versatile Disks, (DVDs), etc.), among others, and transmission type media such as digital and analog communication links.

[0265] One embodiment includes, a method for optimizing search, the method, can  
5 include, one or more of: identifying a location identifier of a webpage having content associated with a semantic type; extracting a pattern from the location identifier of the webpage; storing the pattern in a computer database embodied in a computer-readable storage medium; wherein, the pattern corresponds to a semantic type of the content in the webpage; using the pattern stored on the computer-readable storage medium, performing  
10 pattern recognition on multiple location identifiers having a same domain name as the location identifier; detecting matching location identifiers from the multiple location identifiers having a same pattern as the pattern extracted from the location identifier; identifying, a set of type-determined web pages having the matching location identifiers; wherein, each of the set of type-determined web pages have content that is related to the  
15 semantic type.

[0266] In one embodiment, the location identifier is a Universal Resource Identifier (URI) or Universal Resource Locator (URL). The method can include, generating a set of search results from the set of type-determined web pages. The method can further include, associating each of the set of type-determined web-pages with an ontology defining the  
20 semantic type. The method can further include, extracting additional semantic data from HTML or XHTML content of each of the set of type-determined web pages; and generating a refined set of search results from the set of search results using the additional semantic data. In one embodiment, the additional semantic metadata is extracted using GRDDL (Gleaning Resource Descriptions from Dialects of Languages).

[0267] The method can include, data mining content of the set of type-determined web pages to further determine relevancy to the semantic type; identifying a refined set of type-determined web pages from the set of type-determined web pages based on the relevancy to the semantic type determined via the data mining; and generating a refined set of search results from the refined set of type-determined web pages. The method can include,  
25 ranking each of the set of type-determined web pages based on the relevancy determined from the data mining. In one embodiment, the multiple location identifiers are identified from a sitemap. The method can include, automatically parsing the location identifier to  
30

extract the pattern based on the semantic type. In one embodiment, the pattern is extracted from the location identifier using user input. In one embodiment, the semantic type of the content of the webpage is automatically determined.

5 [0268] In one embodiment, the semantic type of the content of the webpage is defined by an end user. In one embodiment, the semantic type is defined by the end user through tagging or annotating the webpage on a user interface. In one embodiment, the pattern for the semantic type is determined when the webpage is identified as having content of the semantic type by the end user. The method can include, identifying multiple patterns for the semantic type based on identification of multiple webpage as having content of the  
10 semantic type by the end user or additional end users. In one embodiment, the semantic type is associated with multiple attributes that are, one or more of, user-defined and automatically determined. In one embodiment, the multiple attributes are rated by users via a user interface. In one embodiment, the multiple attributes of the semantic type is determined by a predefined ontology or user defined ontology.

15 [0269] In one embodiment, the pattern further corresponds to an attribute of the semantic type of the content in the webpage; and wherein, each of the type-determined web pages have content that is associated with the attribute of the semantic type. In one embodiment, the pattern includes a wildcard segment. In one embodiment, the pattern further includes a domain name segment. In one embodiment, the pattern further includes a semantic type  
20 segment. In one embodiment, the pattern further includes an attributes segment.

[0270] The method can include: extracting a sub-pattern from the location identifier of the webpage; wherein, the sub-pattern corresponds to an attribute of the semantic type of the content.

25 [0271] The method can include: using the sub-pattern, performing pattern recognition on the multiple location identifiers; detecting a subset of the matching location identifiers having a same pattern as the sub-pattern; identifying a subset of the type-determined web pages having the subset of the matching location identifiers; wherein, each of the subset of the type-determined web pages include content associated with the attribute of the semantic type.

30 [0272] The method can include: extracting multiple sub-patterns from the location identifier; wherein, each of the multiple sub-patterns corresponds to one of the multiple

attributes of the semantic type; identifying multiple subsets of the subsets of the type-determined web pages using the multiple sub-patterns; wherein, each of the multiple subsets of the type-determined web pages include content associated with one of the multiple attributes of the semantic type.

- 5 [0273] The method can include: presenting, on a first page of a user interface, the type-determined web pages categorized as having the semantic type; presenting, on the first page of the user interface, links to each of the multiple subsets of the type-determined web pages that are associated with one of the multiple attributes of the semantic type; wherein, each of the links are identified, on the first page, by each of the multiple attributes;
- 10 presenting, on a second page of the user interface, another set of type-determined web pages as categorized as having another semantic type; wherein, each of the links are identified, on the first page, by each of the multiple attributes.

- [0274] The method can include: presenting, a type-determined web page that is hosted by a sponsoring domain, as a search result that has a differentiating template. The method can
- 15 include: performing pattern recognition using the pattern on additional location identifiers having a different domain name than the location identifier; detecting an additional set of matching location identifiers from the additional location identifiers having the same pattern as the pattern extracted from the location identifier; generating an additional set of search results including an additional set of the type-determined web pages having the
- 20 additional set of matching location identifiers; wherein, each of the additional set of type-determined web pages having content associated with the semantic type.

- [0275] A method of search, the method, comprising: detecting a set of location identifiers that have a pattern that matches a specified pattern, the specified pattern being stored in a computer-readable storage medium; wherein, the specified pattern corresponds to a
- 25 semantic type; identifying a set of search results as having content related to the semantic type; wherein, the set of search results includes objects associated with the set of location identifiers having the specified pattern.

- [0276] The method can include, linking the objects to an ontology defining the semantic type. In one embodiment, the location identifier is a Universal Resource Identifier (URI) or
- 30 Universal Resource Locator (URL); wherein, the objects are web pages. The method can include, data mining content of each of the set of search results to further determine

relevancy to the semantic type; and identifying a refined set of search results from the set of type-determined web pages based on the relevancy to the semantic type determine via the data mining.

[0277] The method can include, ranking each of the set of search results based on the relevancy determined from the data mining. In one embodiment, the set of location identifiers are identified from links of user-added content in a social network. In one embodiment, the specified pattern is automatically identified by performing pattern recognition using parameters determined from the semantic type. In one embodiment, the specified pattern is manually specified and reconfigurable.

5 [0278] A method for optimizing search, the method, comprising: identifying a pattern template followed in one or more URIs, the pattern template including multiple specified patterns being useful to resolve a plurality of semantic object types; storing the pattern template in a computer database embodied in a computer-readable storage medium for use in subsequent searches; providing a user interface that enables a user to perform searching;  
10 receiving a search query via the user interface; retrieving the pattern template from the computer database; mapping the search query into a specified pattern of the multiple specified patterns; performing a search for matching URIs having a pattern that matches the specified pattern. The method can include, presenting web pages corresponding to the matching URIs as search results to the user.

20 [0279] A method for optimizing search using semantics, the method, comprising: identifying a Universal Resource Identifier (URI) of a webpage having content associated with a semantic type as specified by a user; identifying a pattern, from the URI of the webpage, that corresponds to the semantic type of the content in the webpage, based on the semantic type that is specified by the user; storing the pattern in a computer-readable  
25 storage medium; identifying additional patterns for the semantic type based on identification of additional webpage as having content of the semantic type by the user; storing the additional patterns in the computer-readable storage medium; using the pattern and the additional patterns, performing pattern recognition on multiple URIs; detecting matching URIs from the multiple URIs having a same pattern as the pattern extracted from  
30 the URI; identifying a set of type-determined web pages having the matching URIs; wherein, each of the set of type-determined web pages have content that is of the semantic type.

[0280] The method can include, ranking each of the additional patterns identified for the semantic type based on a number of users that identified each of the additional web pages as having content of the semantic type. In one embodiment, the semantic type is defined by the user through tagging the webpage on a user interface. The method can include:

5 identifying multiple sub-patterns from the URI using multiple attributes of the semantic type defined by the user or other users. The method can include: identifying multiple subsets of the type-determined web pages using the multiple sub-patterns; wherein, each of the multiple subsets of the type-determined web pages include content associated with one of the multiple attributes of the semantic type.

10 [0281] In one embodiment, the webpage is identified by a second user as having content of a second semantic type that is different from the semantic type; identifying a second pattern, from the URI of the webpage, that corresponds to the second semantic type of the content in the webpage, based on the second semantic type that is specified by the second user; wherein, the second pattern is different from or same as the pattern; using the second  
15 pattern, identifying a second set of type-determined web pages; wherein, each of the second set of type-determined web pages have content that is of the second semantic type; and presenting, to the second user, the second set of type-determined web pages categorized as having the second semantic type; wherein, the set of type-determined web pages are presented to the user as having the semantic type.

20 [0282] A system for optimizing search using semantics, the system, comprising: a URI identification module that identifies a Universal Resource Identifier (URI) of a webpage having content associated with a semantic type; an extraction module that extracts a pattern from the URI of the webpage; wherein, the pattern corresponds to a semantic type of the content in the webpage; a recognition module that performs pattern recognition, using the  
25 pattern on multiple URIs having a same domain name as the URI; wherein, the recognition module detects matching URIs from the multiple URIs having a same pattern as the pattern extracted from the URI; a results generator module that identifies a set of type-determined web pages having the matching URIs; wherein, each of the set of type-determined web pages have content that is related to the semantic type; the extraction module further  
30 extracts a sub-pattern from the URI of the webpage having content associated with the semantic type; wherein, the sub-pattern corresponds to an attribute of the semantic type of the content; the recognition module further performs pattern recognition on the multiple URIs and detects a subset of the matching URIs having a same pattern as the sub-pattern;

the results generator module further identifies a subset of the type-determined web pages having the subset of the matching URIs; wherein, each of the subset of the type-determined web pages include content associated with the attribute of the semantic type.

[0283] The system can further include: a user interface module that presents, on a first page in a user interface, the type-determined web pages categorized as having the semantic type; the user interface module presents, on the first page in the user interface, links to each of the multiple subsets of the type-determined web pages that are associated with one of the multiple attributes of the semantic type; wherein, each of the links are identified, on the first page, by each of the multiple attributes.

[0284] A machine-readable storage medium having stored thereon a set of instructions which when executed perform a method of search, the method, comprising, one or more of: detecting a set of Universal Resource Identifiers (URIs) that have a pattern that matches a specified pattern; wherein, the specified pattern corresponds to a semantic type; identifying a set of search results as having content related to the semantic type from web pages associated with the set of URIs having the specified pattern; data mining content of each of the set of search results to further determine relevancy to the semantic type; identifying a refined set of search results from the set of type-determined web pages based on the relevancy to the semantic type determine via the data mining; wherein, the set of URIs are identified from links of user-added content in a social network.

[0285] A system of a search engine, comprising, one or more of: means for, detecting a set of Universal Resource Identifiers (URIs) that have a pattern that matches a specified pattern; wherein, the specified pattern corresponds to a semantic type; and/or means for, identifying a set of search results as having content related to the semantic type from web pages associated with the set of URIs having the specified pattern.

[0286] Unless the context clearly requires otherwise, throughout the description and the claims, the words "comprise," "comprising," and the like are to be construed in an inclusive sense, as opposed to an exclusive or exhaustive sense; that is to say, in the sense of "including, but not limited to." As used herein, the terms "connected," "coupled," or any variant thereof, means any connection or coupling, either direct or indirect, between two or more elements; the coupling of connection between the elements can be physical, logical, or a combination thereof. Additionally, the words "herein," "above," "below," and

words of similar import, when used in this application, shall refer to this application as a whole and not to any particular portions of this application. Where the context permits, words in the above Detailed Description using the singular or plural number may also include the plural or singular number respectively. The word "or," in reference to a list of two or more items, covers all of the following interpretations of the word: any of the items in the list, all of the items in the list, and any combination of the items in the list.

[0287] The above detailed description of embodiments of the disclosure is not intended to be exhaustive or to limit the teachings to the precise form disclosed above. While specific embodiments of, and examples for, the disclosure are described above for illustrative purposes, various equivalent modifications are possible within the scope of the disclosure, as those skilled in the relevant art will recognize. For example, while processes or blocks are presented in a given order, alternative embodiments may perform routines having steps, or employ systems having blocks, in a different order, and some processes or blocks may be deleted, moved, added, subdivided, combined, and/or modified to provide alternative or subcombinations. Each of these processes or blocks may be implemented in a variety of different ways. Also, while processes or blocks are at times shown as being performed in series, these processes or blocks may instead be performed in parallel, or may be performed at different times. Further any specific numbers noted herein are only examples: alternative implementations may employ differing values or ranges.

[0288] The teachings of the disclosure provided herein can be applied to other systems, not necessarily the system described above. The elements and acts of the various embodiments described above can be combined to provide further embodiments.

[0289] Any patents and applications and other references noted above, including any that may be listed in accompanying filing papers, are incorporated herein by reference.

Aspects of the disclosure can be modified, if necessary, to employ the systems, functions, and concepts of the various references described above to provide yet further embodiments of the disclosure.

[0290] These and other changes can be made to the disclosure in light of the above Detailed Description. While the above description describes certain embodiments of the disclosure, and describes the best mode contemplated, no matter how detailed the above appears in text, the teachings can be practiced in many ways. Details of the system may

vary considerably in its implementation details, while still being encompassed by the subject matter disclosed herein. As noted above, particular terminology used when describing certain features or aspects of the disclosure should not be taken to imply that the terminology is being redefined herein to be restricted to any specific characteristics, features, or aspects of the disclosure with which that terminology is associated. In general, the terms used in the following claims should not be construed to limit the disclosure to the specific embodiments disclosed in the specification, unless the above Detailed Description section explicitly defines such terms. Accordingly, the actual scope of the disclosure encompasses not only the disclosed embodiments, but also all equivalent ways of practicing or implementing the disclosure under the claims.

**[0291]** While certain aspects of the disclosure are presented below in certain claim forms, the inventors contemplate the various aspects of the disclosure in any number of claim forms. For example, while only one aspect of the disclosure is recited as a means-plus-function claim under 35 U.S.C. §112, ¶6, other aspects may likewise be embodied as a means-plus-function claim, or in other forms, such as being embodied in a computer-readable medium. (Any claims intended to be treated under 35 U.S.C. §112, ¶6 will begin with the words “means for”.) Accordingly, the applicant reserves the right to add additional claims after filing the application to pursue such additional claim forms for other aspects of the disclosure.



Claims

*What is claimed is:*

- 5     1.     A method for optimizing search, the method, comprising:  
            identifying a location identifier of a webpage having content associated  
            with a semantic type;  
            extracting a pattern from the location identifier of the webpage;  
            storing the pattern in a computer database embodied in a computer-readable  
10     storage medium;  
            wherein, the pattern corresponds to a semantic type of the content in the  
            webpage;  
            using the pattern stored on the computer-readable storage medium,  
            performing pattern recognition on multiple location identifiers having a same  
15     domain name as the location identifier;  
            detecting matching location identifiers from the multiple location identifiers  
            having a same pattern as the pattern extracted from the location identifier;  
            identifying, a set of type-determined web pages having the matching  
            location identifiers;  
20     wherein, each of the set of type-determined web pages have content that is  
            related to the semantic type.
- 25     2.     The method of claim 1, wherein, the location identifier is a Universal Resource  
            Identifier (URI) or Universal Resource Locator (URL).
- 30     3.     The method of claim 1, further comprising, generating a set of search results from  
            the set of type-determined web pages.
4.     The method of claim 1, further comprising, associating each of the set of type-  
            determined web-pages with an ontology defining the semantic type.

5. The method of claim 1, further comprising,  
extracting additional semantic data from HTML or XHTML content of each  
of the set of type-determined web pages; and  
generating a refined set of search results from the set of search results using  
5 the additional semantic data.
6. The method of claim 5, wherein, the additional semantic metadata is extracted  
using GRDDL (Gleaning Resource Descriptions from Dialects of Languages).
- 10 7. The method of claim 3, further comprising,  
data mining content of the set of type-determined web pages to further  
determine relevancy to the semantic type;  
identifying a refined set of type-determined web pages from the set of type-  
determined web pages based on the relevancy to the semantic type determined via  
15 the data mining; and  
generating a refined set of search results from the refined set of type-  
determined web pages.
8. The method of claim 7, further comprising, ranking each of the set of type-  
20 determined web pages based on the relevancy determined from the data mining.
9. The method of claim 1, wherein, the multiple location identifiers are identified  
from a sitemap.
- 25 10. The method of claim 1, further comprising, automatically parsing the location  
identifier to extract the pattern based on the semantic type.
11. The method of claim 1, wherein, the pattern is extracted from the location identifier  
using user input.

30

12. The method of claim 1, wherein, the semantic type of the content of the webpage is automatically determined.
13. The method of claim 1, wherein, the semantic type of the content of the webpage is defined by an end user.
14. The method of claim 13, wherein, the semantic type is defined by the end user through tagging or annotating the webpage on a user interface.
15. The method of claim 13, wherein, the pattern for the semantic type is determined when the webpage is identified as having content of the semantic type by the end user.
16. The method of claim 13, further comprising, identifying multiple patterns for the semantic type based on identification of multiple webpage as having content of the semantic type by the end user or additional end users.
17. The method of claim 1, wherein, the semantic type is associated with multiple attributes that are, one or more of, user-defined and automatically determined.
18. The method of claim 17, wherein, the multiple attributes are rated by users via a user interface.
19. The method of claim 17, wherein, the multiple attributes of the semantic type is determined by a predefined ontology or user defined ontology.
20. The method of claim 1,  
wherein, the pattern further corresponds to an attribute of the semantic type of the content in the webpage; and  
wherein, each of the type-determined web pages have content that is associated with the attribute of the semantic type.
21. The method of claim 1, wherein, the pattern includes a wildcard segment.

22. The method of claim 21, wherein, the pattern further includes a domain name segment.
23. The method of claim 21, wherein, the pattern further includes a semantic type segment.
24. The method of claim 21, wherein, the pattern further includes an attributes segment.
25. The method of claim 1, further comprising:  
extracting a sub-pattern from the location identifier of the webpage;  
wherein, the sub-pattern corresponds to an attribute of the semantic type of the content.
26. The method of claim 25, further comprising:  
using the sub-pattern, performing pattern recognition on the multiple location identifiers;  
detecting a subset of the matching location identifiers having a same pattern as the sub-pattern;  
identifying a subset of the type-determined web pages having the subset of the matching location identifiers;  
wherein, each of the subset of the type-determined web pages include content associated with the attribute of the semantic type.
27. The method of claim 1, further comprising:  
extracting multiple sub-patterns from the location identifier;  
wherein, each of the multiple sub-patterns corresponds to one of the multiple attributes of the semantic type;  
identifying multiple subsets of the subsets of the type-determined web pages using the multiple sub-patterns;  
wherein, each of the multiple subsets of the type-determined web pages include content associated with one of the multiple attributes of the semantic type.

28. The method of claim 27, further comprising:

presenting, on a first page of a user interface, the type-determined web  
5 pages categorized as having the semantic type;

presenting, on the first page of the user interface, links to each of the  
multiple subsets of the type-determined web pages that are associated with one of  
the multiple attributes of the semantic type;

wherein, each of the links are identified, on the first page, by each of the  
10 multiple attributes;

presenting, on a second page of the user interface, another set of type-  
determined web pages as categorized as having another semantic type;

wherein, each of the links are identified, on the first page, by each of the  
multiple attributes.

29. The method of claim 1, further comprising:

presenting, a type-determined web page that is hosted by a sponsoring  
domain, as a search result that has a differentiating template.

30. The method of claim 1, further comprising:

performing pattern recognition using the pattern on additional location  
identifiers having a different domain name than the location identifier;

detecting an additional set of matching location identifiers from the  
additional location identifiers having the same pattern as the pattern extracted from  
25 the location identifier;

generating an additional set of search results including an additional set of  
the type-determined web pages having the additional set of matching location  
identifiers;

wherein, each of the additional set of type-determined web pages having  
30 content associated with the semantic type.

31. A method of search, the method, comprising:  
detecting a set of location identifiers that have a pattern that matches a  
specified pattern, the specified pattern being stored in a computer-readable storage  
medium;  
5 wherein, the specified pattern corresponds to a semantic type;  
identifying a set of search results as having content related to the semantic  
type;  
wherein, the set of search results includes objects associated with the set of  
location identifiers having the specified pattern.
- 10 32. The method of claim 31, further comprising, linking the objects to an ontology  
defining the semantic type.
33. The method of claim 31,  
15 wherein, the location identifier is a Universal Resource Identifier (URI) or  
Universal Resource Locator (URL);  
wherein, the objects are web pages.
34. The method of claim 33, further comprising,  
20 data mining content of each of the set of search results to further determine  
relevancy to the semantic type; and  
identifying a refined set of search results from the set of type-determined  
web pages based on the relevancy to the semantic type determine via the data  
mining.
- 25 35. The method of claim 33, further comprising, ranking each of the set of search  
results based on the relevancy determined from the data mining.
36. The method of claim 31, wherein, the set of location identifiers are identified from  
30 links of user-added content in a social network.
37. The method of claim 31, wherein, the specified pattern is automatically identified  
by performing pattern recognition using parameters determined from the semantic  
type.

38. The method of claim 31, wherein, the specified pattern is manually specified and reconfigurable.

5 39. A method for optimizing search, the method, comprising:

identifying a pattern template followed in one or more URIs, the pattern template including multiple specified patterns being useful to resolve a plurality of semantic object types;

10 storing the pattern template in a computer database embodied in a computer-readable storage medium for use in subsequent searches;

providing a user interface that enables a user to perform searching;

receiving a search query via the user interface;

retrieving the pattern template from the computer database;

15 mapping the search query into a specified pattern of the multiple specified patterns;

performing a search for matching URIs having a pattern that matches the specified pattern.

20 40. The method of claim 62, further comprising, presenting web pages corresponding to the matching URIs as search results to the user.

41. A method for optimizing search using semantics, the method, comprising:

identifying a Universal Resource Identifier (URI) of a webpage having content associated with a semantic type as specified by a user;

25 identifying a pattern, from the URI of the webpage, that corresponds to the semantic type of the content in the webpage, based on the semantic type that is specified by the user;

storing the pattern in a computer-readable storage medium;

30 identifying additional patterns for the semantic type based on identification of additional webpage as having content of the semantic type by the user;

storing the additional patterns in the computer-readable storage medium;

using the pattern and the additional patterns, performing pattern recognition on multiple URIs;

detecting matching URIs from the multiple URIs having a same pattern as the pattern extracted from the URI;

identifying a set of type-determined web pages having the matching URIs;

wherein, each of the set of type-determined web pages have content that is of the semantic type.

42. The method of claim 41, further comprising, ranking each of the additional patterns identified for the semantic type based on a number of users that identified each of the additional web pages as having content of the semantic type.

43. The method of claim 41, wherein, the semantic type is defined by the user through tagging the webpage on a user interface.

44. The method of claim 35, further comprising:

identifying multiple sub-patterns from the URI using multiple attributes of the semantic type defined by the user or other users.

45. The method of claim 42, further comprising:

identifying multiple subsets of the type-determined web pages using the multiple sub-patterns;

wherein, each of the multiple subsets of the type-determined web pages include content associated with one of the multiple attributes of the semantic type.

46. The method of claim 41,

wherein, the webpage is identified by a second user as having content of a second semantic type that is different from the semantic type;

identifying a second pattern, from the URI of the webpage, that corresponds to the second semantic type of the content in the webpage, based on the second semantic type that is specified by the second user;

wherein, the second pattern is different from or same as the pattern;  
using the second pattern, identifying a second set of type-determined web pages;

wherein, each of the second set of type-determined web pages have content that is of the second semantic type; and



presenting, to the second user, the second set of type-determined web pages categorized as having the second semantic type;

wherein, the set of type-determined web pages are presented to the user as having the semantic type.

5

47. A system for optimizing search using semantics, the system, comprising:

a URI identification module that identifies a Universal Resource Identifier (URI) of a webpage having content associated with a semantic type;

an extraction module that extracts a pattern from the URI of the webpage;

10

wherein, the pattern corresponds to a semantic type of the content in the webpage;

a recognition module that performs pattern recognition, using the pattern on multiple URIs having a same domain name as the URI;

15

wherein, the recognition module detects matching URIs from the multiple URIs having a same pattern as the pattern extracted from the URI;

a results generator module that identifies a set of type-determined web pages having the matching URIs;

wherein, each of the set of type-determined web pages have content that is related to the semantic type;

20

the extraction module further extracts a sub-pattern from the URI of the webpage having content associated with the semantic type;

wherein, the sub-pattern corresponds to an attribute of the semantic type of the content;

25

the recognition module further performs pattern recognition on the multiple URIs and detects a subset of the matching URIs having a same pattern as the sub-pattern;

the results generator module further identifies a subset of the type-determined web pages having the subset of the matching URIs;

30

wherein, each of the subset of the type-determined web pages include content associated with the attribute of the semantic type.

48. The system of claim 47, further comprising:

a user interface module that presents, on a first page in a user interface, the type-determined web pages categorized as having the semantic type;

the user interface module presents, on the first page in the user interface, links to each of the multiple subsets of the type-determined web pages that are associated with one of the multiple attributes of the semantic type;

wherein, each of the links are identified, on the first page, by each of the multiple attributes.

49. A machine-readable storage medium having stored thereon a set of instructions which when executed perform a method of search, the method, comprising:

detecting a set of Universal Resource Identifiers (URIs) that have a pattern that matches a specified pattern;

wherein, the specified pattern corresponds to a semantic type;

identifying a set of search results as having content related to the semantic type from web pages associated with the set of URIs having the specified pattern;

data mining content of each of the set of search results to further determine relevancy to the semantic type;

identifying a refined set of search results from the set of type-determined web pages based on the relevancy to the semantic type determine via the data mining;

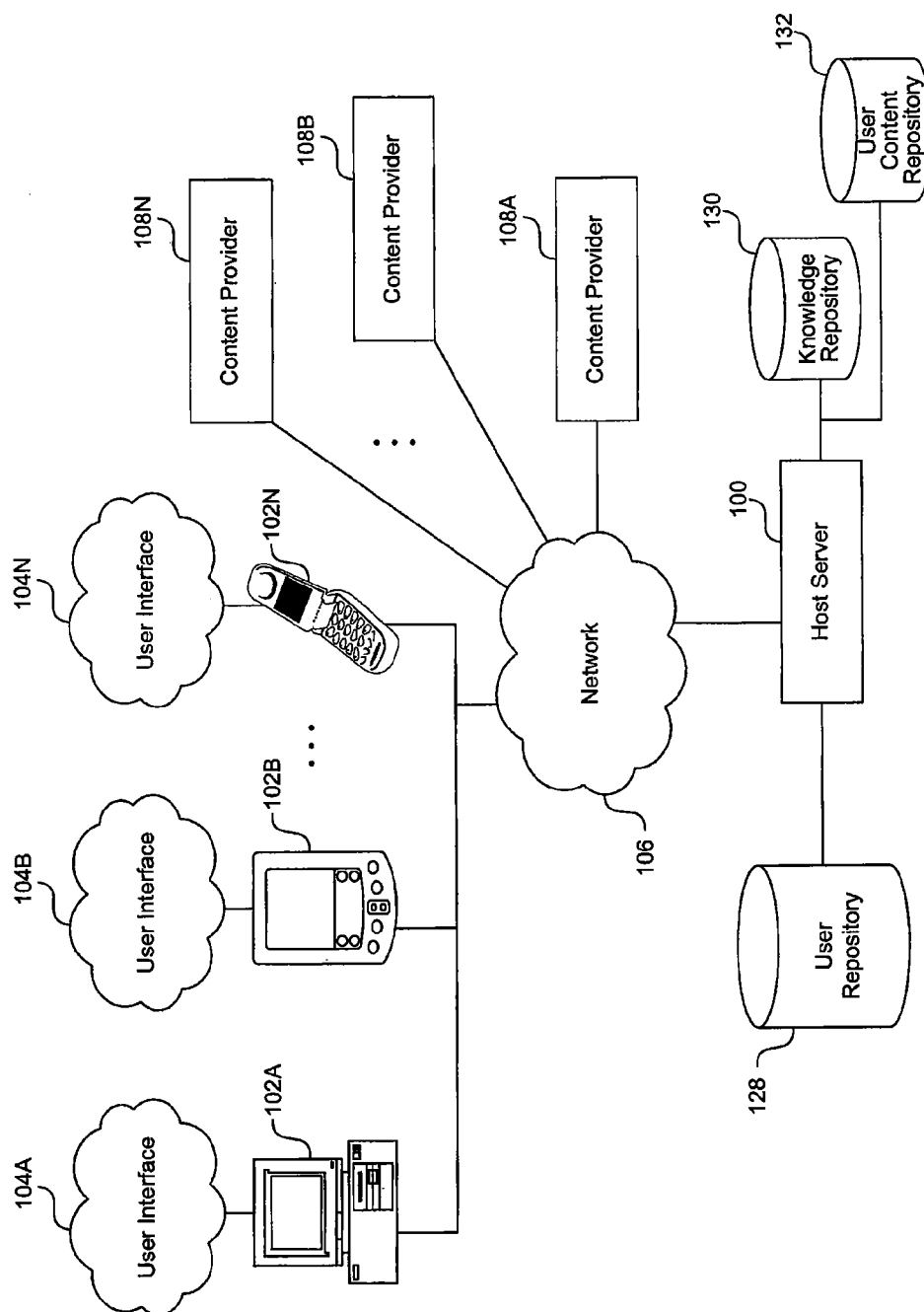
wherein, the set of URIs are identified from links of user-added content in a social network.

50. A system of a search engine, comprising:

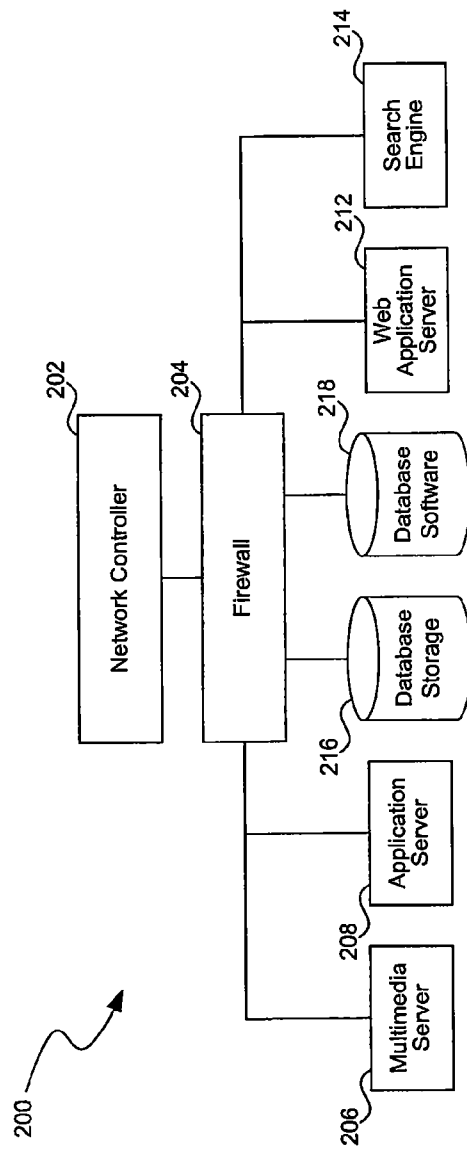
means for, detecting a set of Universal Resource Identifiers (URIs) that have a pattern that matches a specified pattern;

wherein, the specified pattern corresponds to a semantic type;

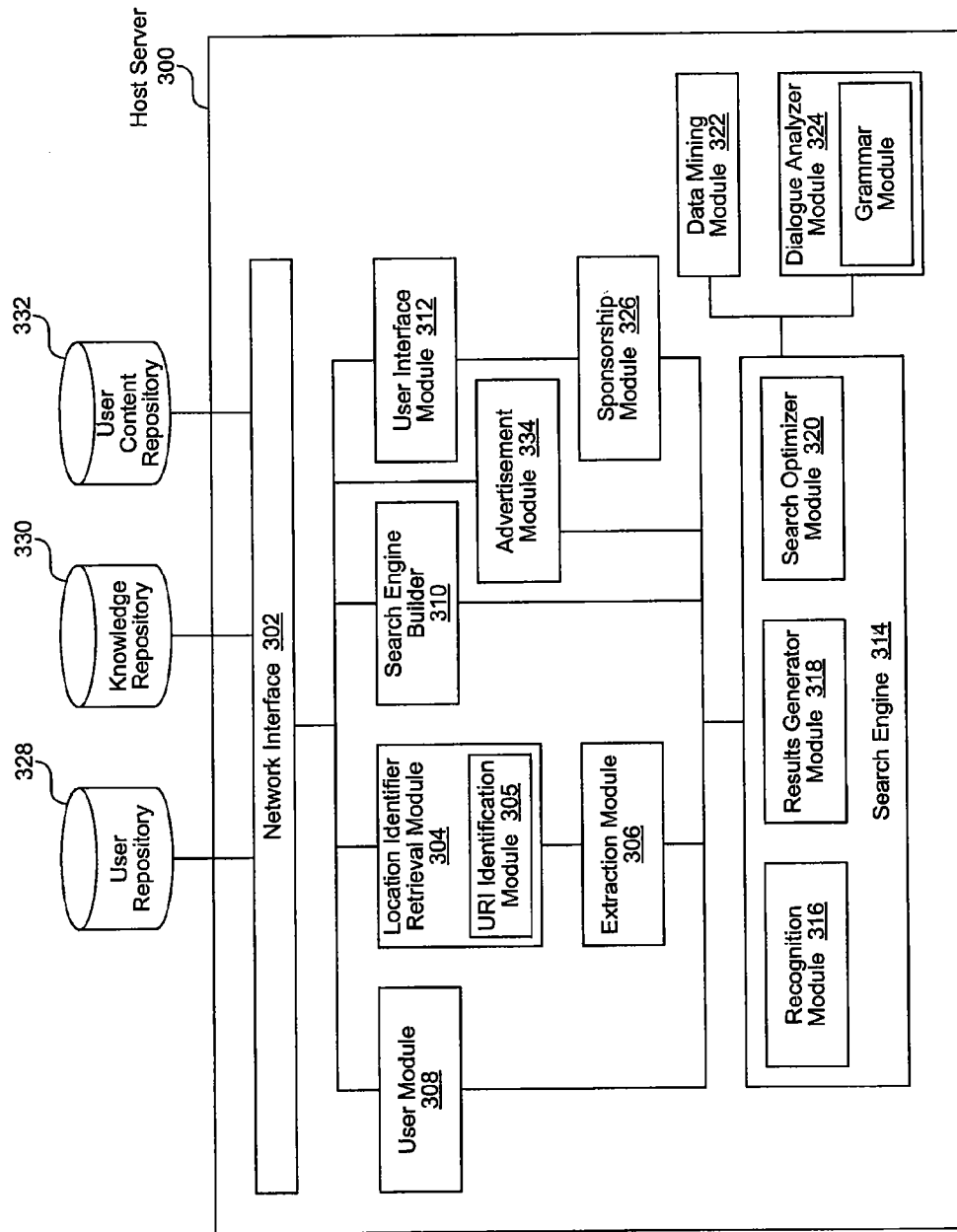
means for, identifying a set of search results as having content related to the semantic type from web pages associated with the set of URIs having the specified pattern.



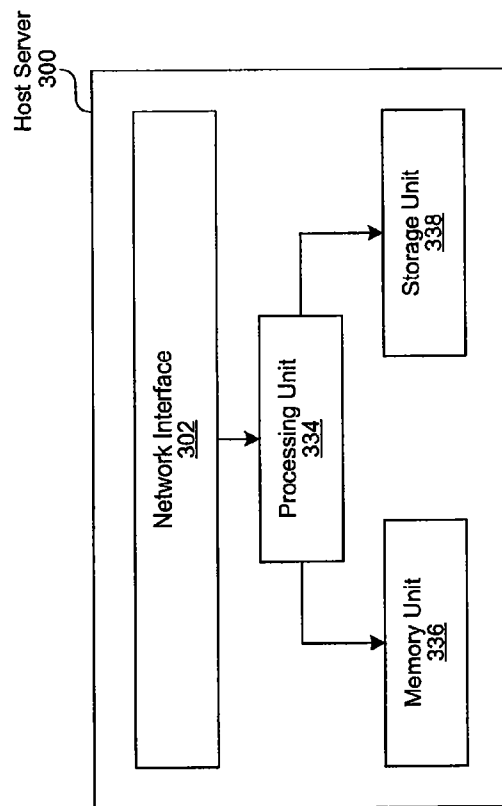
**FIG. 1**

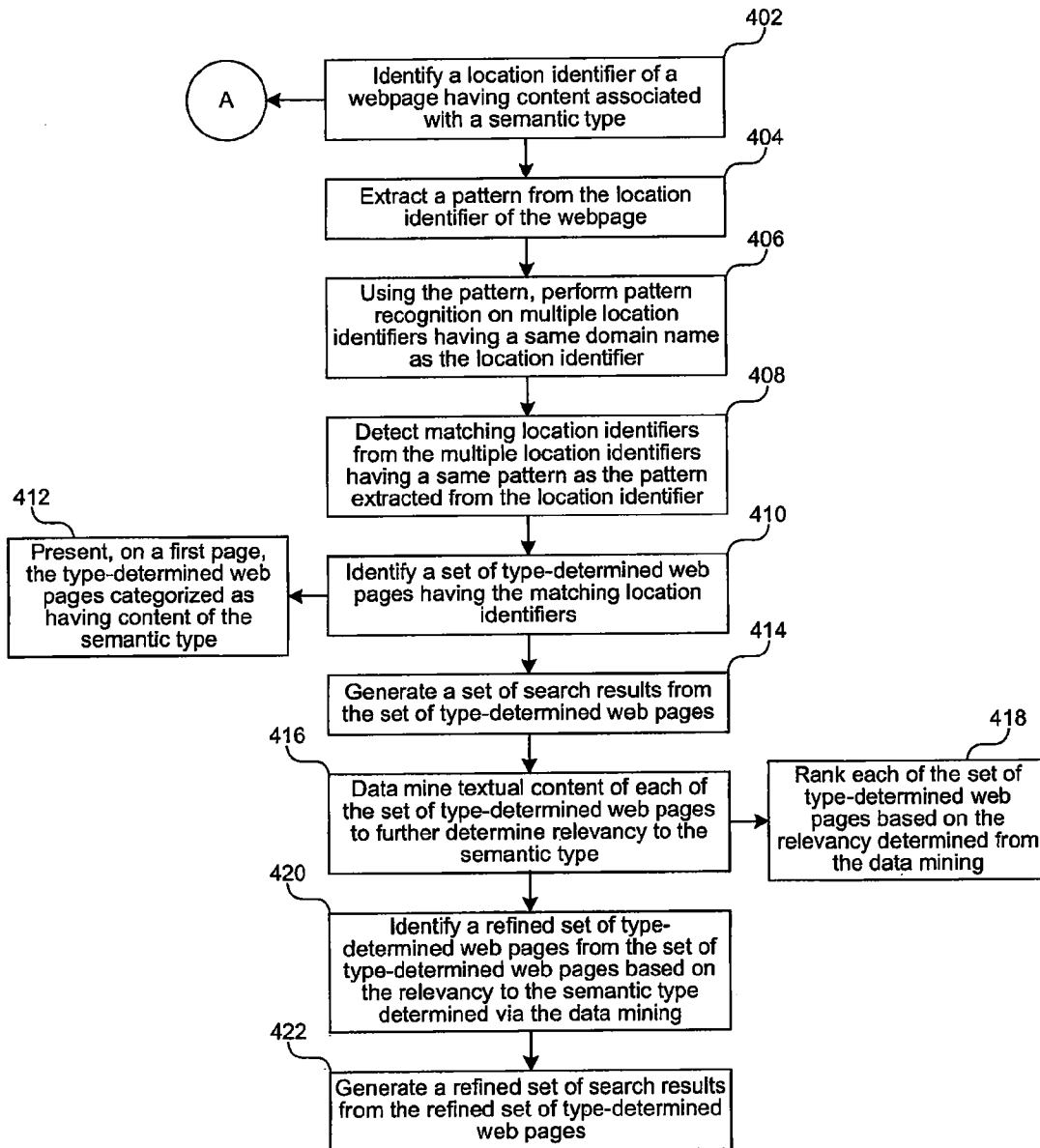


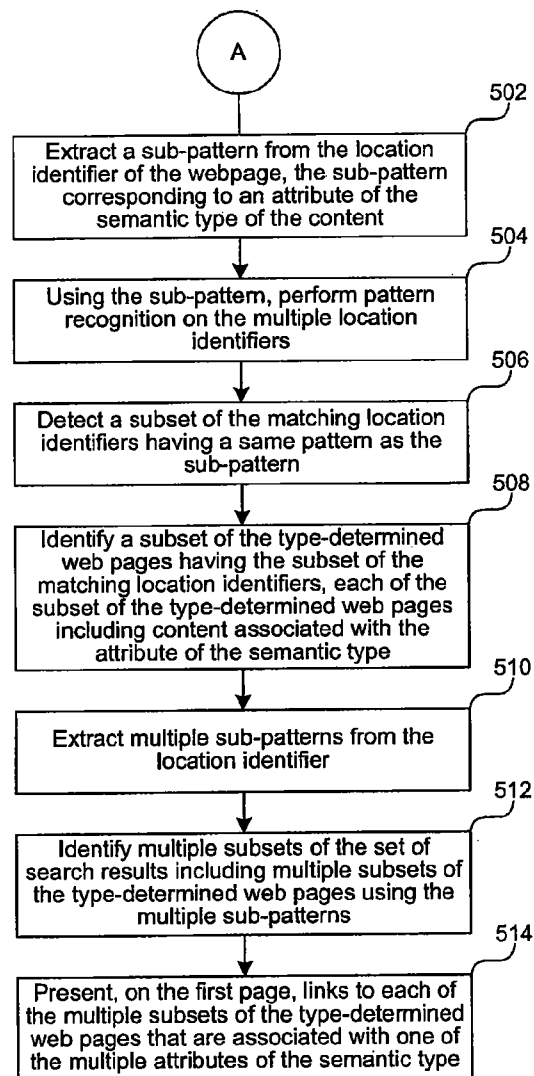
**FIG. 2**



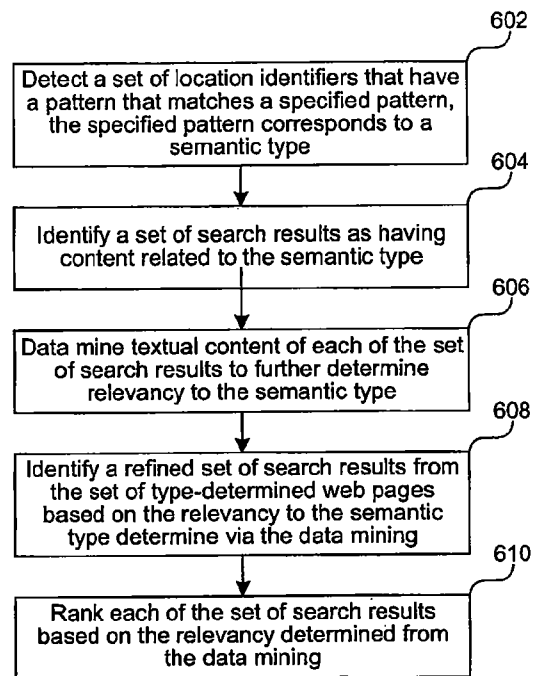
**FIG. 3A**

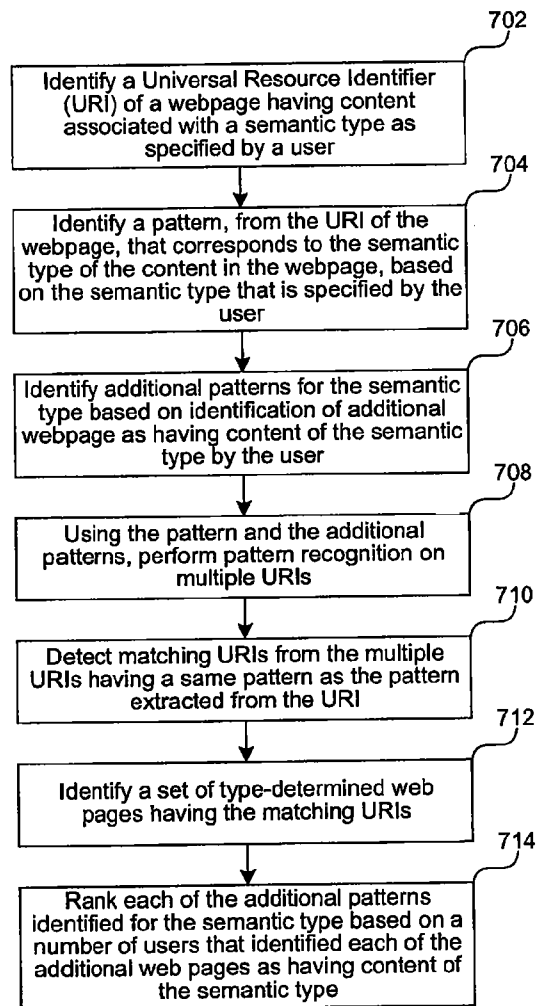
**FIG. 3B**

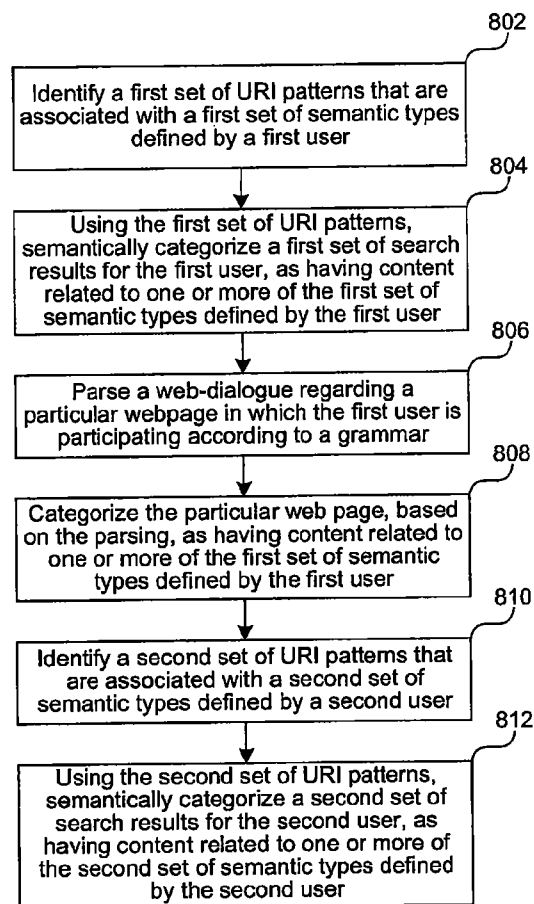
**FIG. 4**

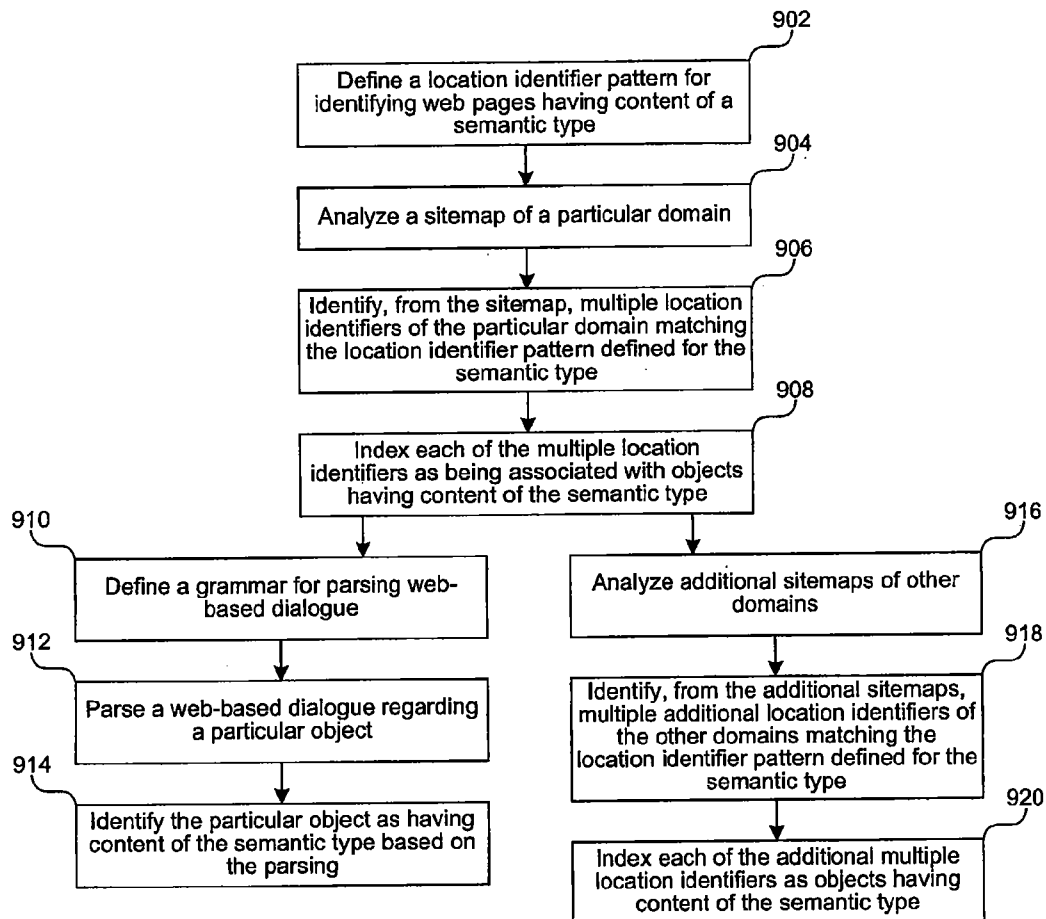
**FIG. 5**

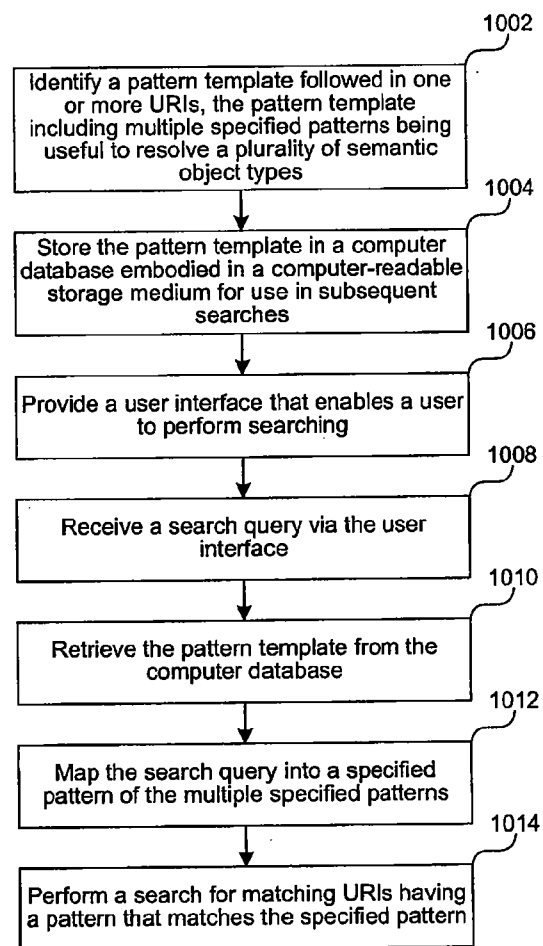


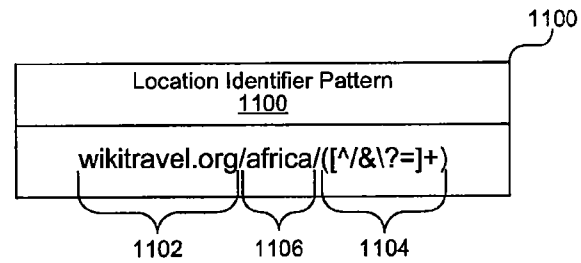
**FIG. 6**

**FIG. 7**

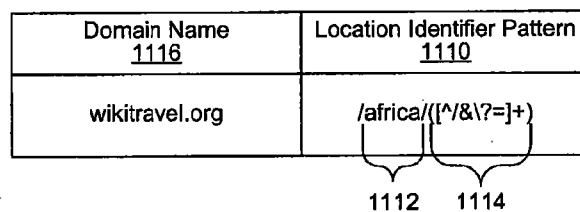
**FIG. 8**

**FIG. 9**

**FIG. 10**



**FIG. 11A**



**FIG. 11B**

twine

Start >> Domains >> wikitravel.org

**Domain: wikitravel.org**

**Details**

Name	wikitravel.org
Partnership Level	<input type="text" value="None"/>
Sitemap Status	<input type="text" value="Run"/>
Sitemap URLs	1600
Last Sitemap Scan	Mar 31, 2009 at 12:10 PM

**Mapped URL Patterns**

domain	type	pattern	user	dateAdded
wikitravel.org	44	<input type="text" value=".../wiki/en/index.php?title=([^\&amp;\?]=)history"/>	john	2009-03-31
wikitravel.org	44	<input type="text" value=".../africa/([^\&amp;\?]=)"/>	john	2009-03-31

**Templates**

type
<input type="text" value="Travel"/>
<input type="text" value="History"/>
<input type="text" value="Define New"/>

**Web Resources**

[Edit Stylesheet](#)

name	key
File: <input type="text"/>	<input type="button" value="Browse..."/> <input type="button" value="Upload"/>

**FIG. 12A**

twine   Search

[Start >> Domains >> wikitravel.org](#)

### Pattern for Travel at wikitravel.org

---

#### Details

Domain	wikitravel.org
Pattern	<input type="text" value="http://wikitravel.org/en/([^\&amp;\?=\ ]+)"/> 1204
Type	Travel
Added By	john
Date Added	Mar 31, 2009 at 12:18 PM
Validated	true

---

#### Statistics

 1212

---

Matching Pages 1355

---

#### Pattern Group Mappings

 1214

---

Group #	Property	Transform	
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="button" value="Add"/>
1216	1218	1220	

---

#### Static Mappings

 1222

---

Property	Value	
<input type="text"/>	<input type="text"/>	<input type="button" value="Add"/>
1224	1226	

---

#### Knowledge Transform

 1228

---

[Edit Transform](#)

**FIG. 12B**

1225



twine

[Start >> Domains >> wikitravel.org](#)

### Pattern for History at wikitravel.org

---

**Details**

Domain	<a href="#">wikitravel.org</a>
Pattern	<a href="#">http://wikitravel.org/wiki/en/index.php?title=([^\&amp;\?]=)=history</a> 1206
Type	<a href="#">History</a>
Added By	john
Date Added	Mar 31, 2009 at 12:10 PM
Validated	true

---

**Statistics** 1230

Matching Pages 136

---

**Pattern Group Mappings** 1232

Group #	Property	Transform	
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="button" value="Add"/>

---

**Static Mappings** 1234

Property	Value	
<input type="text"/>	<input type="text"/>	<input type="button" value="Add"/>

---

**Knowledge Transform** 1236

[Edit Transform](#)

**FIG. 12C**

1250

twine

Start >> Domains >> wikitravel.org

**Domain: www.mayoclinic.com** 1302

**Details**

Name	www.mayoclinic.com		
Partnership Level	<input type="text" value="Sponsored Content"/> 1312	<input type="text" value="None"/> <input type="text" value="Sponsored Content"/> <input type="text" value="Hosted Content"/>	
Sitemap Status	<input type="text" value="Run"/>		
Sitemap URLs	<input type="text" value="26650"/>		
Last Sitemap Scan	<input type="text" value="Feb 20, 2009 at 09:00 PM"/>		

**Mapped URL Patterns**

domain	type	pattern	user	dateAdded
www.mayoclinic.com	14	<input type="text" value=".../health/drug-information(['^&amp;?=?=]+)"/> 1304	john	2009-02-24
www.mayoclinic.com	11	<input type="text" value=".../health/(['^&amp;?=?=]+)/DS(['^&amp;?=?=]+)"/> 1306	john	2009-02-22

**Templates**

type	
Disease	<input type="text" value=""/> 1308
New	<input type="text" value=""/> 1310

**Web Resources** 1314

[Edit Stylesheet](#)

name	key	
mayo-titlebg.gif	www.mayoclinic.com/web/resources/mayo-titlebg.gif	<input type="text" value="(X)"/>
logo_mayoclinic.gif	www.mayoclinic.com/web/resources/logo_mayoclinic.gif	<input type="text" value="(X)"/>
title_mayoclinic.gif	www.mayoclinic.com/web/resources/title_mayoclinic.gif	<input type="text" value="(X)"/>
mayo-logobg.png	www.mayoclinic.com/web/resources/mayo-logobg.png	<input type="text" value="(X)"/>
mayo-blurbbg.png	www.mayoclinic.com/web/resources/mayo-blurbbg.png	<input type="text" value="(X)"/>

File:

FIG. 13A

1300

twine

Search

Search

Domains

Vocabulary

Admin

Start >> Domains >> wikitravel.org

Pattern for Drug at www.mayoclinic.com

Details

[Edit]

Domain

www.mayoclinic.com

Pattern

http://www1.mayoclinic.com/health/drug-information/DS([^\&?=\+] 1304

Type

Drug

Added By

john

Date Added

Feb 24, 2009 at 07:58 PM

Validated

true

Statistics

Matching Pages 0

Pattern Group Mappings

Group #

Property

Transform

Add

Static Mappings

Property

Value

Add

Knowledge Transform

Edit Transform

FIG. 13B

1325



**Pattern for Disease at [www.mayoclinic.com](http://www.mayoclinic.com)**

## Details

**[Edit]**

Domain	<a href="http://www.mayoclinic.com">www.mayoclinic.com</a>
Pattern	<a href="http://www.mayoclinic.com/health/&lt;[^\&amp;?=-]+/DS([^\&amp;?=-]+)">http://www.mayoclinic.com/health/&lt;[^\&amp;?=-]+/DS([^\&amp;?=-]+)</a>
Type	<u>Disease</u>
Added By	john
Date Added	Feb 22, 2009 at 01:26 PM
Validated	true

## Statistics

Matching Pages 0

## Pattern Group Mappings

Group #	Property	Transform	
1	title	fromSlug()	Remove
			Add

## Static Mappings

Property	Value	Add

## Knowledge Transform

### Edit Transform

**FIG. 13C**

~ 1350



twine   Search

Start >> Domains >> wikitravel.org

Type:

Template  1406 1408

```
.sponsored {  
}  
.sponsored table {  
border-collapse: collapse;  
}  
.sponsored ul.sections li {  
display: block;  
/* white-space: nowrap; */  
line-height: 90%;  
padding: 0 4px 8px 5px;  
margin-right: 5px;  
/* border-right: 1px solid #d7e1e5; */  
}  
.sponsored ul.sections li a {  
color: #144f68;  
}  
.sponsored ul.sections li:before {  
content: '';  
}  
.sponsored p {  
}
```

Preview Resource:

**FIG. 14B**

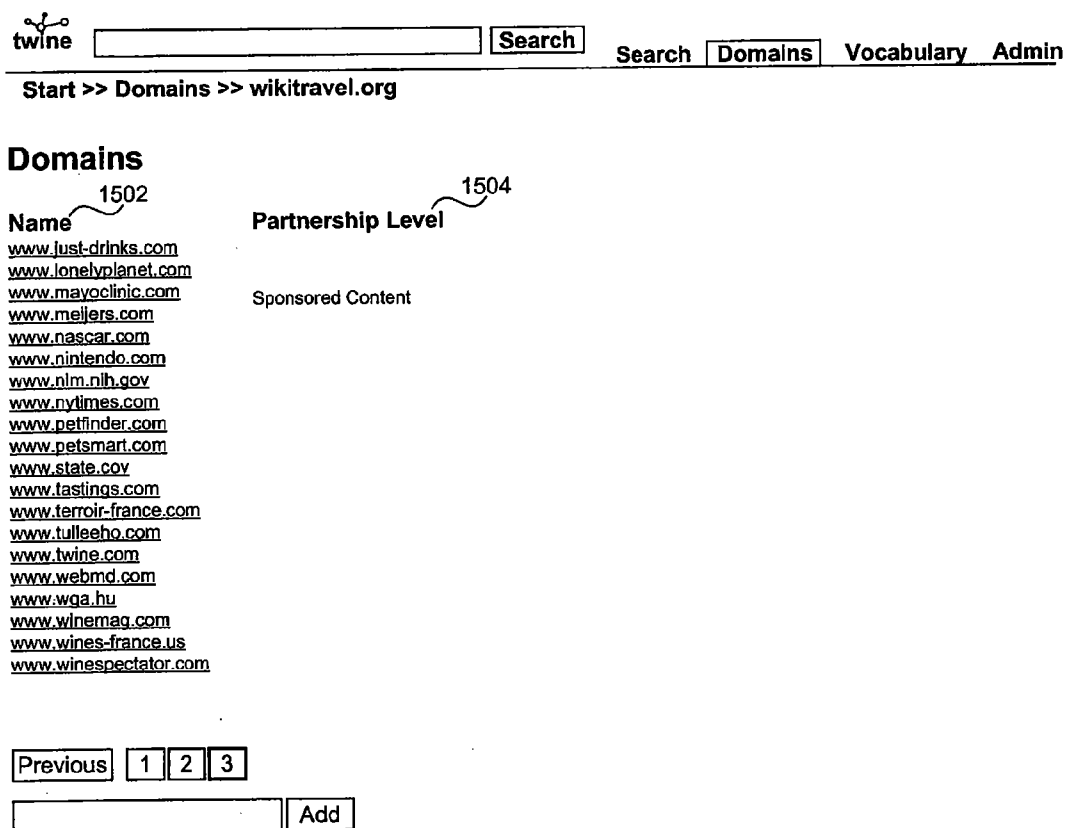


FIG. 15

1500

twine       1601

Start >> Domains >> wikitravel.org

Vocabulary

1602

**Types** | **Properties**

Name	Added	Added By	Validated
Article	Feb 20, 2009	john	false
Artist	Feb 24, 2009	john	false
Automobile Manufacturer	Feb 21, 2009	john	false
Automobile Model	Feb 28, 2009	john	false
Automobile Model Review	Feb 11, 2009	john	false
Automobile News Article	Feb 11, 2009	john	false
Beverage Industry	Mar 31, 2009	john	false
Beverages	Mar 31, 2009	john	false
Bordeaux Wines	Mar 31, 2009	john	false
Burgundy Wine	Mar 31, 2009	john	false
Business News	Mar 31, 2009	john	false
City	Feb 20, 2009	john	false
Country	Mar 26, 2009	john	false
Dance	Mar 26, 2009	john	false
Dance Magazine	Mar 26, 2009	john	false
Dance Type	Mar 26, 2009	john	false
Disease	Feb 22, 2009	john	false
Dog Breed	Feb 23, 2009	john	false
Drug	Feb 24, 2009	john	false

1604

1 2 3 | Next

FIG. 16A

1600



twine   Search Domains  Admin

Start >> Domains >> wikitravel.org

Vocabulary

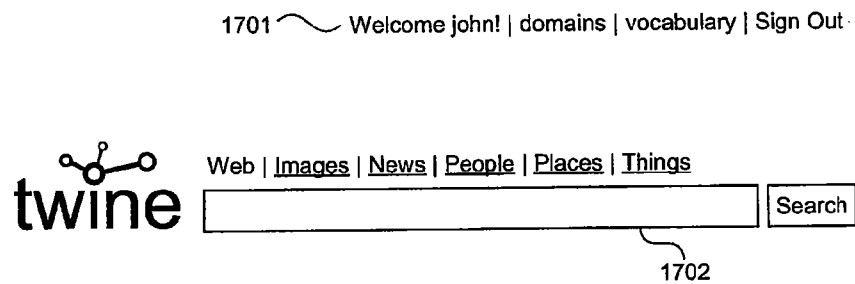
Types  1606

Name	Datatype	Search Facet	Added	Added By	Validated
Available Engine	String		Feb 28, 2009	john	false
Available Engine Type	String		Feb 28, 2009	john	false
Available Transmission	String		Feb 28, 2009	john	false
Breed Group	String	yes	Feb 23, 2009	john	false
Capacity	String		Mar 01, 2009	john	false
City	String		Mar 01, 2009	john	false
City Mpg / Lower	String		Feb 28, 2009	john	false
City Mpg / Upper	String		Feb 28, 2009	john	false
Comfort Feature	String	yes	Mar 01, 2009	john	false
Consumer Ratings Range	String	yes	Mar 01, 2009	john	false
Highway Mpg / Lower	String		Feb 28, 2009	john	false
Highway Mpg / Upper	String		Mar 28, 2009	john	false
Horsepower Range	String	yes	Mar 01, 2009	john	false
Information Source	String	yes	Feb 25, 2009	john	false
Lower Mpg	String		Mar 01, 2009	john	false
Lower Msrp	String		Feb 28, 2009	john	false
Make	String	yes	Feb 21, 2009	john	false
Model Name	String	yes	Feb 21, 2009	john	false
Model Type	String	yes	Mar 01, 2009	john	false
Model Year	String	yes	Feb 21, 2009	john	false
Price	String		Mar 01, 2009	john	false

1608 {

| Next

FIG. 16B



**FIG. 17**

Welcome john! | domains | vocabulary | Sign Out

twine fever Search

All Images News Diseases (52) Drugs (15) Medical Conditions (9) Products (4)

Showing 1 to 5 of 9327297 total matching results.

1801

1810

**Fever** **Disease**

**Definition | Symptoms | Causes | When to seek medical advice | Tests and diagnosis | Complications | Treatments and drugs | Prevention | Lifestyles and home remedies**

A fever isn't an illness itself, but it's usually a sign that something out of the ordinary is going on in your body. Fevers aren't necessarily bad. In fact, fevers seem to play a key role in helping your body fight off a number of infections.

<http://www.mayoclinic.com/health/fever/DS00077>

Review This | Share This | Discuss This | Extract Pattern

1850

1802 1804 1806 1808

**Related Searches**

fever  
Mayo Clinic  
fever in children  
Research  
Scarlet fever  
Q fever

**Website**

[www.mayoclinic.com](http://www.mayoclinic.com)  
(52)  
[www.drugs.com](http://www.drugs.com) (24)  
[www.gnc.com](http://www.gnc.com) (4)

**Fever - Wikipedia**  
User-generated article about fever, also known as pyrexia, a febrile response for unnatural changes within the body.  
Related: fever, set-point, body temperature, thermoregulatory, body, ...  
[en.wikipedia.org/wiki/Fever](http://en.wikipedia.org/wiki/Fever) - 84k - 2009/03/29  
What's This? | Share This | Discuss This | Extract Pattern

**Fever - MedicineNet**  
Provides information about fever, causes, symptoms, measuring methods, and treatment for one of the most misunderstood symptoms in all medicine.  
Related: fever, thermometer, MedicineNet, high fever, body temperature, ...  
[www.mediclne.com/aches\\_pain\\_fever/article.htm](http://www.mediclne.com/aches_pain_fever/article.htm) - 56k - 2009/03/30  
What's This? | Share This | Discuss This | Extract Pattern

**Fever: First Aid - MayoClinic**  
Offers first-aid steps to take if you or someone else has a fever including how to take a temperature.  
Related: fever, Mayo Clinic, First aid, thermometer, rectal temperature, ...  
[www.mayoclinic.com/health/first-aid-fever/FA00063](http://www.mayoclinic.com/health/first-aid-fever/FA00063) - 29k - 2009/03/30  
What's This? | Share This | Discuss This | Extract Pattern

**Fever - MedlinePlus Encyclopedia**  
General overview of the condition when a person's body temperature exceeds normal, including common causes and home care.  
Related: fever, high fever, normal body temperature, fever in adults, fever in children, ...  
[en.wikipedia.org/wiki/Fever](http://en.wikipedia.org/wiki/Fever) - 84k - 2009/03/29  
What's This? | Share This | Discuss This | Extract Pattern

1 2 3 4 5 | Next

twine

fever

Search

FIG. 18

1800

twine

1901

fever

Search

Welcome john! | domains | vocabulary | Sign Out

All

Images

News

Diseases

Drugs (15)

Medical Conditions (9)

Products (4)

Showing 1 to 3 of 52 total matching results.

1950

**Viral Hemorrhagic Fevers**

**Definition | Symptoms | Causes | Risk factors | Complications | Tests and diagnosis | Treatments and drugs | Prevention**

Viral hemorrhagic fevers (VHFs) are caused by viruses from four distinct families and range in severity from relatively mild to life-threatening. Most viral hemorrhagic fevers are rare in the United States. When they do occur, they're often found in people who've recently traveled internationally.

Although all begin with fever and muscle aches, some viral hemorrhagic fevers progress to more serious problems, including severe internal and external bleeding (hemorrhage), widespread tissue death (necrosis), and shock.

<http://www.mayoclinic.com/health/viral-hemorrhagic-fevers/DS00539>

Review This | Share This | Discuss This | Extract Pattern

1902 1904 1906 1908

**Scarlet Fever**

**Definition | Symptoms | Causes | Risk factors | When to seek medical advice | Tests and diagnosis | Complications | Treatments and drugs | Prevention | Lifestyle and home remedies**

Scarlet fever, or scarlatina, is an illness that brings on a rash covering most of the body, a strawberry-like appearance of the tongue and usually a high fever. The most common source of scarlet fever is one form of a common bacterial infection known as strep throat. Therefore, a sore throat and other signs and symptoms of a typical strep throat infection almost always accompany scarlet fever.

Infrequent sources of scarlet fever are related bacterial infections affecting other organ systems.

<http://www.mayoclinic.com/health/scarlet-fever/DS00917>

Review This | Share This | Discuss This | Extract Pattern

**Hay Fever**

**Definition | Symptoms | Causes | Risk factors | When to seek medical advice | Tests and diagnosis | Complications | Treatments and drugs | Prevention | Alternative medicine**

Hay fever, also called allergic rhinitis, causes cold-like symptoms such as a runny nose, congestion, sneezing and sinus pressure. But unlike a cold, hay fever isn't caused by a virus – it's caused by an allergic response to indoor or outdoor airborne allergens, such as pollen, dust mites or pet dander. Some people have hay fever year-round. For others, hay fever gets worse at certain times of the year, usually in the spring, summer or fall. One of the most common allergic conditions, hay fever affects about one in five people.

<http://www.mayoclinic.com/health/hay-fever/DS00174>

Review This | Share This | Discuss This | Extract Pattern

1910

**Disease**

Ad 1

"Healthcare for Babies"

1922

Ad 2

"Encyclopedia of Fevers"

1924

1 2 3 4 5 6 | Next

twine

fever

Search

FIG. 19A

twine   Welcome john! | domains | vocabulary | Sign Out

All Images News **Diseases** Drugs (15) Medical Conditions (9) Products (4)

Showing 1 to 2 of 52 total matching results.

<b>Definition</b>	<b>Viral Hemorrhagic Fevers</b>	<b>Disease</b>
<b>Symptoms</b>	Viral hemorrhagic fevers (VHFs) are caused by viruses from four distinct families and range in severity from relatively mild to life-threatening. Most viral hemorrhagic fevers are rare in the United States. When they do occur, they're often found in people who've recently traveled internationally.	
<b>Causes</b>		
<b>Risk factors</b>		
<b>Complications</b>	Although all begin with fever and muscle aches, some viral hemorrhagic fevers progress to more serious problems, including severe internal and external bleeding (hemorrhage), widespread tissue death (necrosis), and shock.	
<b>Tests and diagnosis</b>		
<b>Treatments and drugs</b>	No current treatment can cure hemorrhagic fevers. Immunizations exist for only two of the many viral hemorrhagic fevers. Until additional vaccines are developed, the best approach is prevention.	
<b>Prevention</b>		

**Scarlet Fever** **Disease**

**Definition | Symptoms | Causes | Risk factors | When to seek medical advice | Tests and diagnosis | Complications | Treatments and drugs | Prevention | Lifestyle and home remedies**

Scarlet fever, or scarlatina, is an illness that brings on a rash covering most of the body, a strawberry-like appearance of the tongue and usually a high fever. The most common source of scarlet fever is one form of a common bacterial infection known as strep throat. Therefore, a sore throat and other signs and symptoms of a typical strep throat infection almost always accompany scarlet fever.

Infrequent sources of scarlet fever are related bacterial infections affecting other organ systems.

Scarlet fever rarely affects people older than 18 and is most common in children 5 to 15 years of age. Although scarlet fever was once considered a serious childhood illness, antibiotic treatments have made it less threatening.

Nonetheless, if left untreated, scarlet fever (like strep throat) can result in more serious conditions that affect the heart, kidneys and other parts of the body.

<http://www.mayoclinic.com/health/scarlet-fever/DS00917>

[Review This](#) | [Share This](#) | [Discuss This](#) | [Extract Pattern](#)

1

2

3

4

5

6

| Next

twine

**FIG. 19B**

2001

Welcome john! | domains | vocabulary | Sign Out

twine

All Images News Diseases (52) Drugs (15) Medical Conditions (9) Products (4)

Showing 1 to 4 of 1472761 total matching results.

2050 { **Fever Ray**  
 Artist: Fever Ray Album: Fever Ray Genre: Ambient, Electronica Sounds Like: The Knife, Air France, Junior Boys 90 Decibels Fever Ray and its new self-titled album is the work of Karin Dreijer Andersson, one half of Swedish electronic duo The Knife. On her own, Andersson creates sounds and vocals richer than expected from a solo act and extremely telling of trends to come.  
 Source: Daily Orange Posted: 2009/03/31 @ 08:41:07  
 |  |  |

2002 2008

**Hay-Band Acupressure Band Puts Hay Fever Symptoms Under Pressure**  
 LONDON, March 31 /PRNewswire/ — More than 13 million(1) people in Britain suffer from hay fever and five million1 of these could be taking inappropriate medicines, which can affect both their health and hip pocket, however, new developments in acupressure may have the answer in the shape of Hay-Band®.  
 Source: ITNews Posted: 2009/03/31 @ 05:11:18  
 |  |  |

**Medical alert while fever suspect tested**  
 THE emergency unit at Life St Dominic Hospital in East London was closed to the public for several hours on Saturday evening when a patient exhibiting symptoms of a deadly fever was admitted.  
 Source: Daily Dispatch Posted: 2009/03/31 @ 08:45:49  
 |  |  |

**Nigeria: Lassa Fever Caused By Black Rat -Ogunleye**  
 Dr. Olayinka Ogunleye is a Consultant Physician at the Lagos State University Teaching Hospital (LASUTH), Ikeja. He prescribed environmental as well as personal hygienic therapy as a way of keeping Lassa fever at bay, in this interview with Snr Correspondent, Emma Maduabuchi.  
 Source: AllAfrica.com Posted: 2009/03/31 @ 10:51:43  
 |  |  |

**Website**  
[www.mayoclinic.com](http://www.mayoclinic.com)  
 (52)  
[www.drugs.com](http://www.drugs.com) (24)  
[www.gnc.com](http://www.gnc.com) (4)

1 2 3 4 5 6 7 8 | Next

twine

FIG. 20

2000

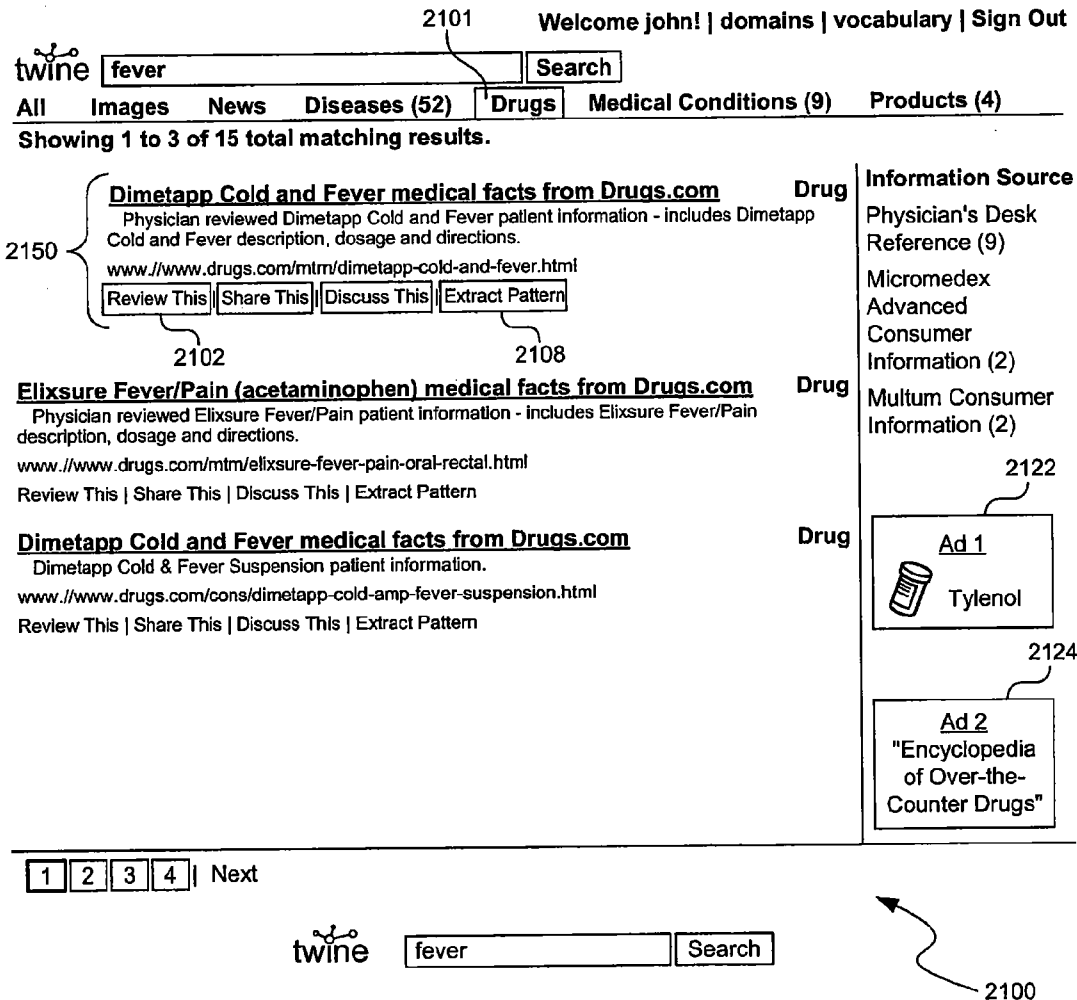


FIG. 21

Welcome john! | domains | vocabulary | Sign Out

twine fever Search 2201

All Images News Diseases (52) Drugs (15) Medical Conditions Products (4)

2250 { **Fever**  
[www.drugs.com/condition/fever.html](http://www.drugs.com/condition/fever.html)  
 Review This | Share This | Discuss This | Extract Pattern  
 2202 2204 2206 2208

**Medical Condition** 2210

**Salmonella Enteric Fever**  
[www.drugs.com/condition/salmonella-enteric-fever.html](http://www.drugs.com/condition/salmonella-enteric-fever.html)  
 Review This | Share This | Discuss This | Extract Pattern  
**Medical Condition**

**Hemorrhagic Fever**  
[www.drugs.com/condition/hemorrhagic-fever.html](http://www.drugs.com/condition/hemorrhagic-fever.html)  
 Review This | Share This | Discuss This | Extract Pattern  
**Medical Condition**

**Familial Mediterranean Fever**  
[www.drugs.com/condition/familial-mediterranean-fever.html](http://www.drugs.com/condition/familial-mediterranean-fever.html)  
 Review This | Share This | Discuss This | Extract Pattern  
**Medical Condition**

**Rheumatic Fever**  
[www.drugs.com/condition/rheumatic-fever.html](http://www.drugs.com/condition/rheumatic-fever.html)  
 Review This | Share This | Discuss This | Extract Pattern  
**Medical Condition**

**Typhoid Fever**  
[www.drugs.com/condition/typhoid-fever.html](http://www.drugs.com/condition/typhoid-fever.html)  
 Review This | Share This | Discuss This | Extract Pattern  
**Medical Condition**

**Rheumatic Fever Prophylaxis**  
[www.drugs.com/condition/rheumatic-fever-prophylaxis.html](http://www.drugs.com/condition/rheumatic-fever-prophylaxis.html)  
 Review This | Share This | Discuss This | Extract Pattern  
**Medical Condition**

1

twine

fever

Search

FIG. 22

2200



Welcome john! | domains | vocabulary | Sign Out

twine fever Search

All Images News Diseases (52) Drugs (15) Medical Conditions (9) Products

2310 Product

2350 {

**Basic Organics L-Lysine Ointment - BASIC ORGANICS - GNC**  
 Shop at GNC for Basic Organics L-Lysine Ointment – BASIC ORGANICS  
<http://www.gnc.com/product/index.jsp?productId=2133725>  
 Review This | Share This | Discuss This | Extract Pattern

2302 2304 2306 2308

**GNC Bio—Remedy® Cold & Flu Relief - BIO REMEDY - GNC**  
 Shop at GNC for GNC Bio—Remedy® Cold & Flu Relief - BIO REMEDY  
<http://www.gnc.com/product/index.jsp?productId=2133750>  
 Review This | Share This | Discuss This | Extract Pattern

**Afrin® No Drip - MILLBROOK 1009163 - GNC**  
 Shop at GNC for Afrin® No Drip - MILLBROOK 1009163  
<http://www.gnc.com/product/index.jsp?productId=2890714>  
 Review This | Share This | Discuss This | Extract Pattern

**GNC Bio—Remedy ♦ Cold & Flue Relief - BIO REMEDY - GNC**  
 Shop at GNC Bio—Remedy ♦ Cold & Flue Relief - BIO REMEDY  
<http://www.gnc.com/product/index.jsp?productId=3190389>  
 Review This | Share This | Discuss This | Extract Pattern

Product

2322 Ad 1 Tylenol

2324 Ad 2 Robitussin AM/PM

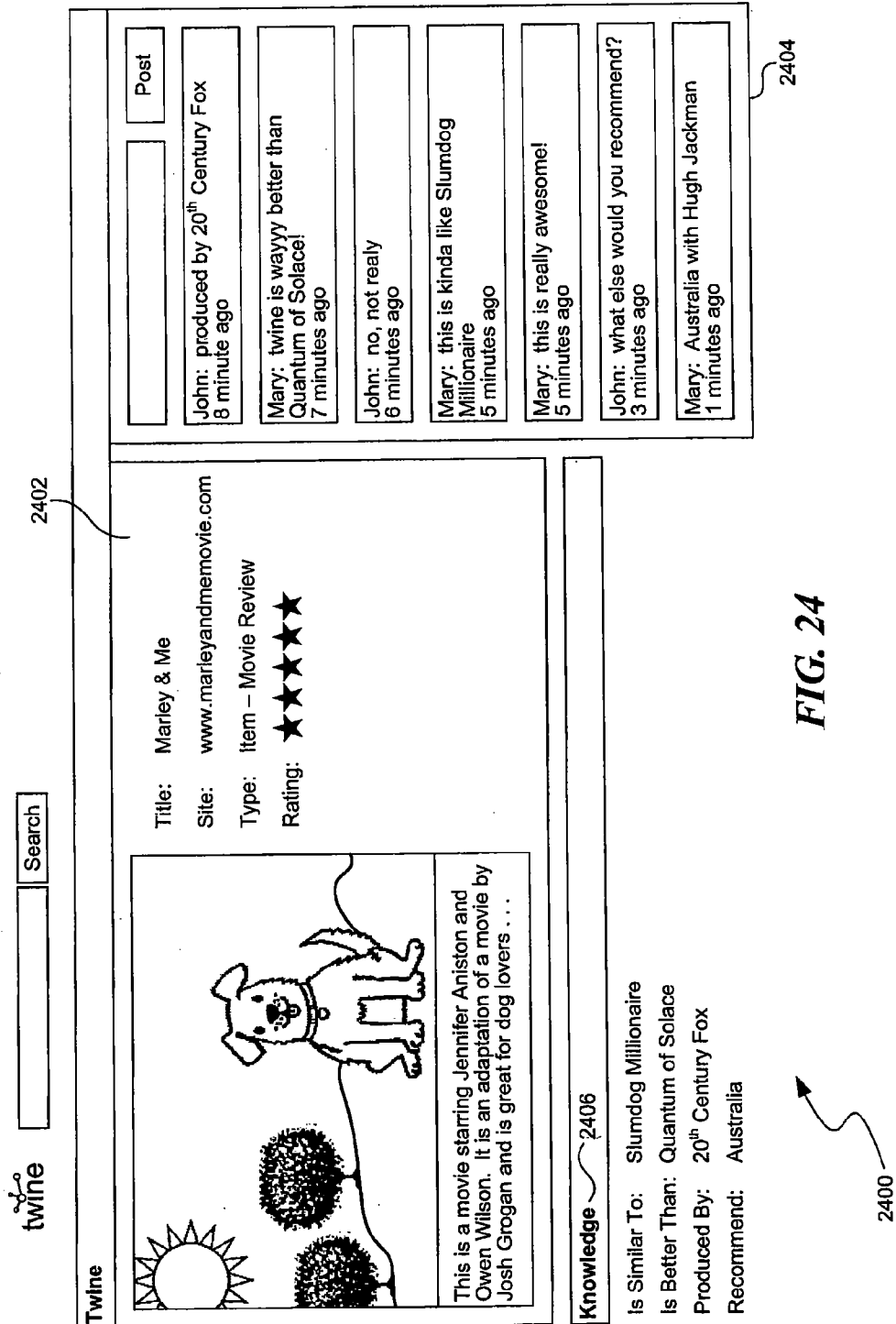
2326 Ad 3 Aspirin

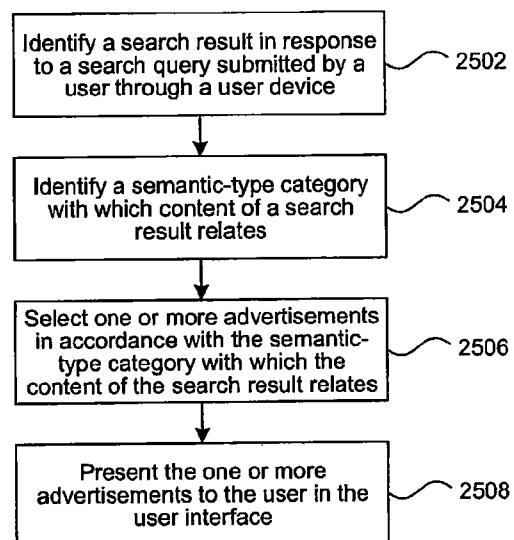
1

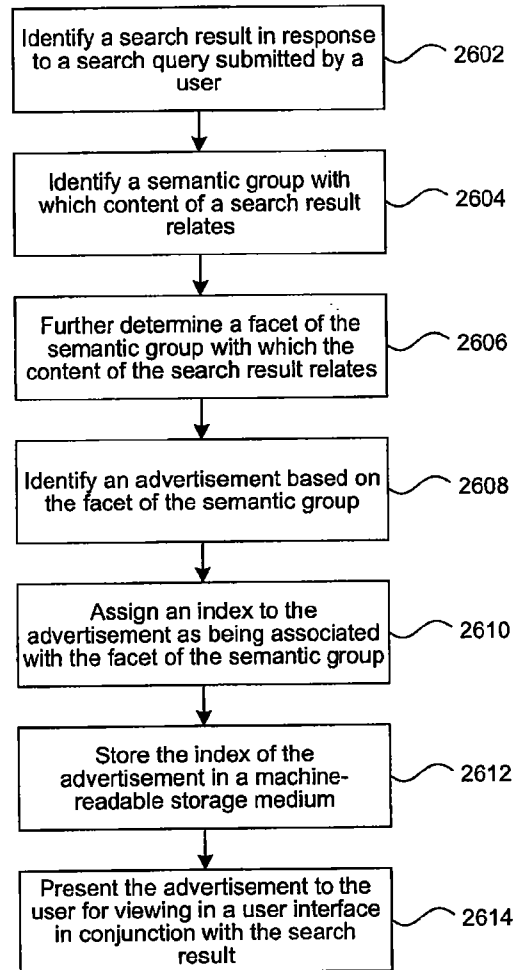
twine fever Search

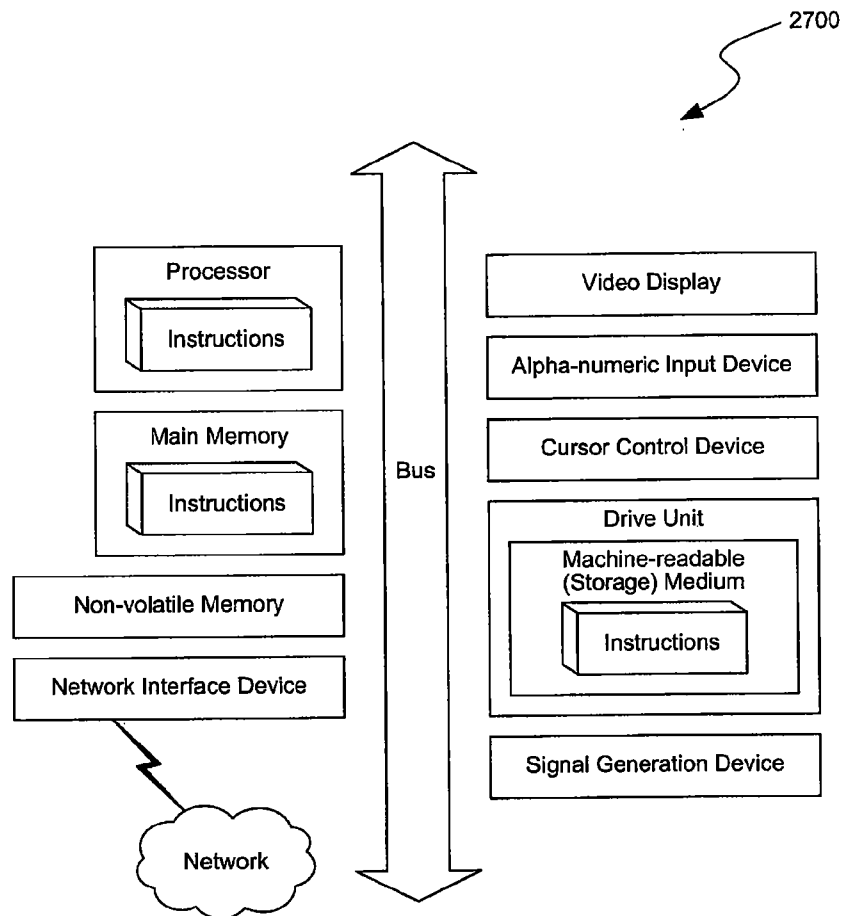
FIG. 23

2300



**FIG. 25**

**FIG. 26**

**FIG. 27**