



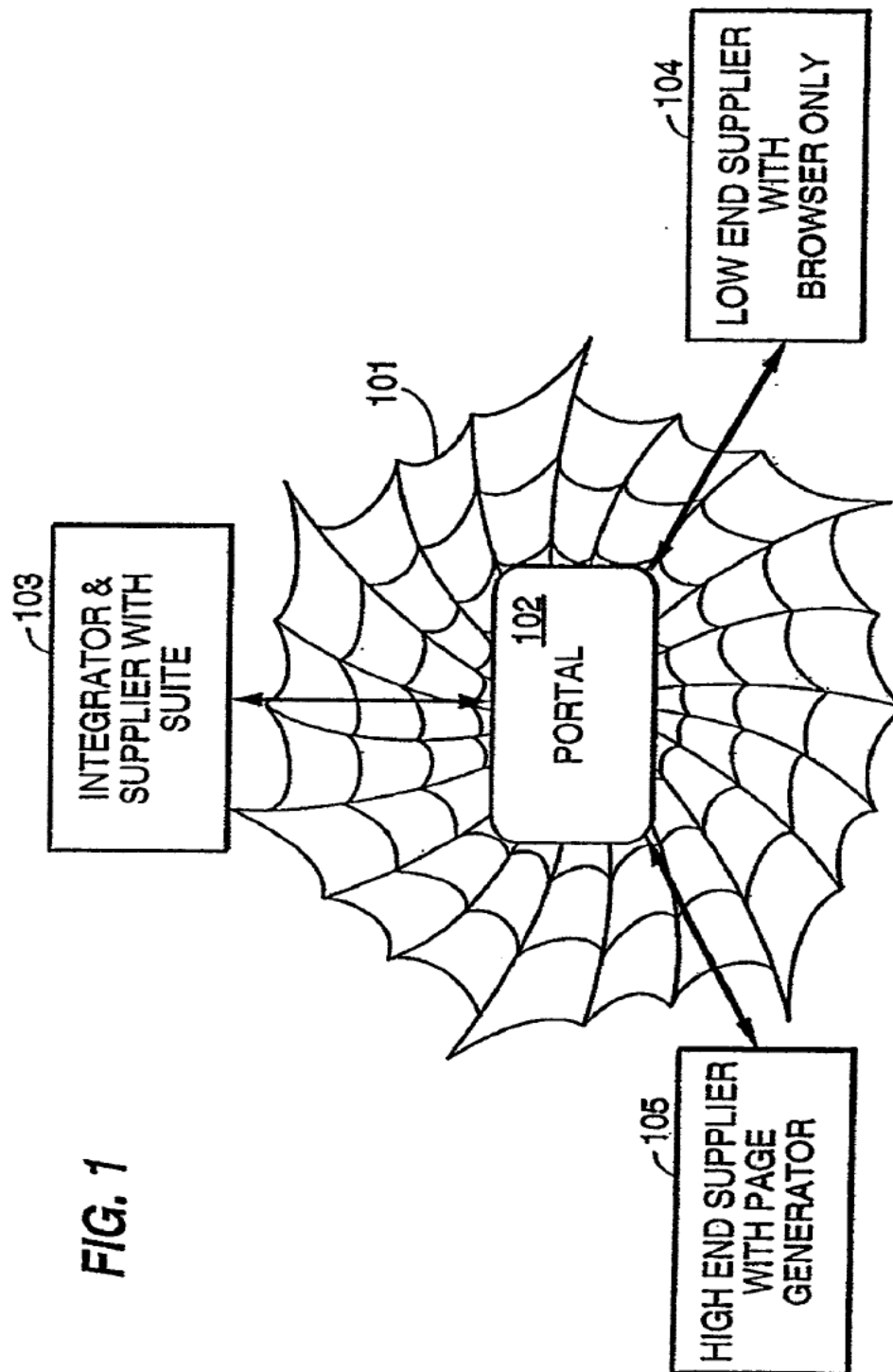
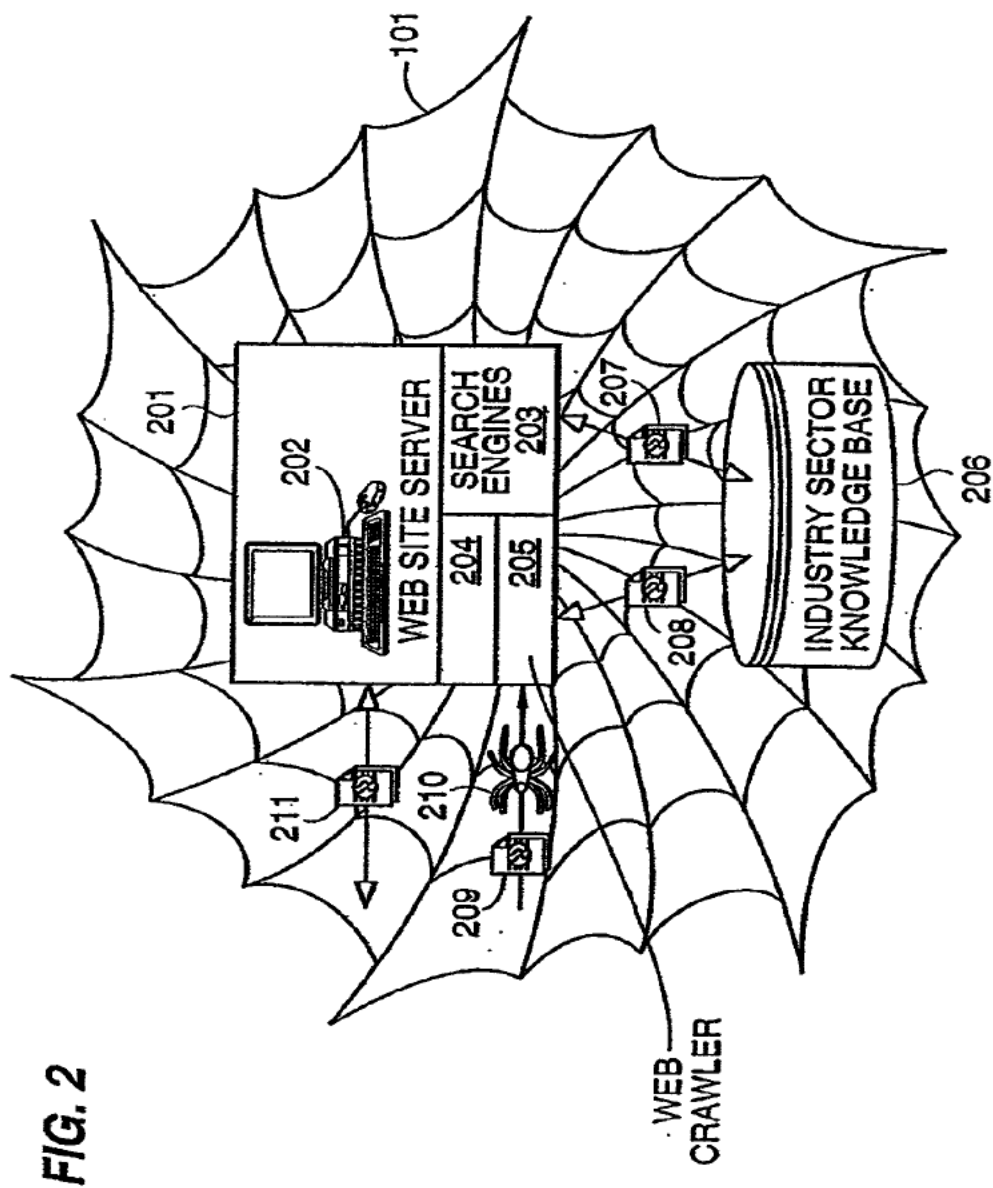


**FIG. 1**



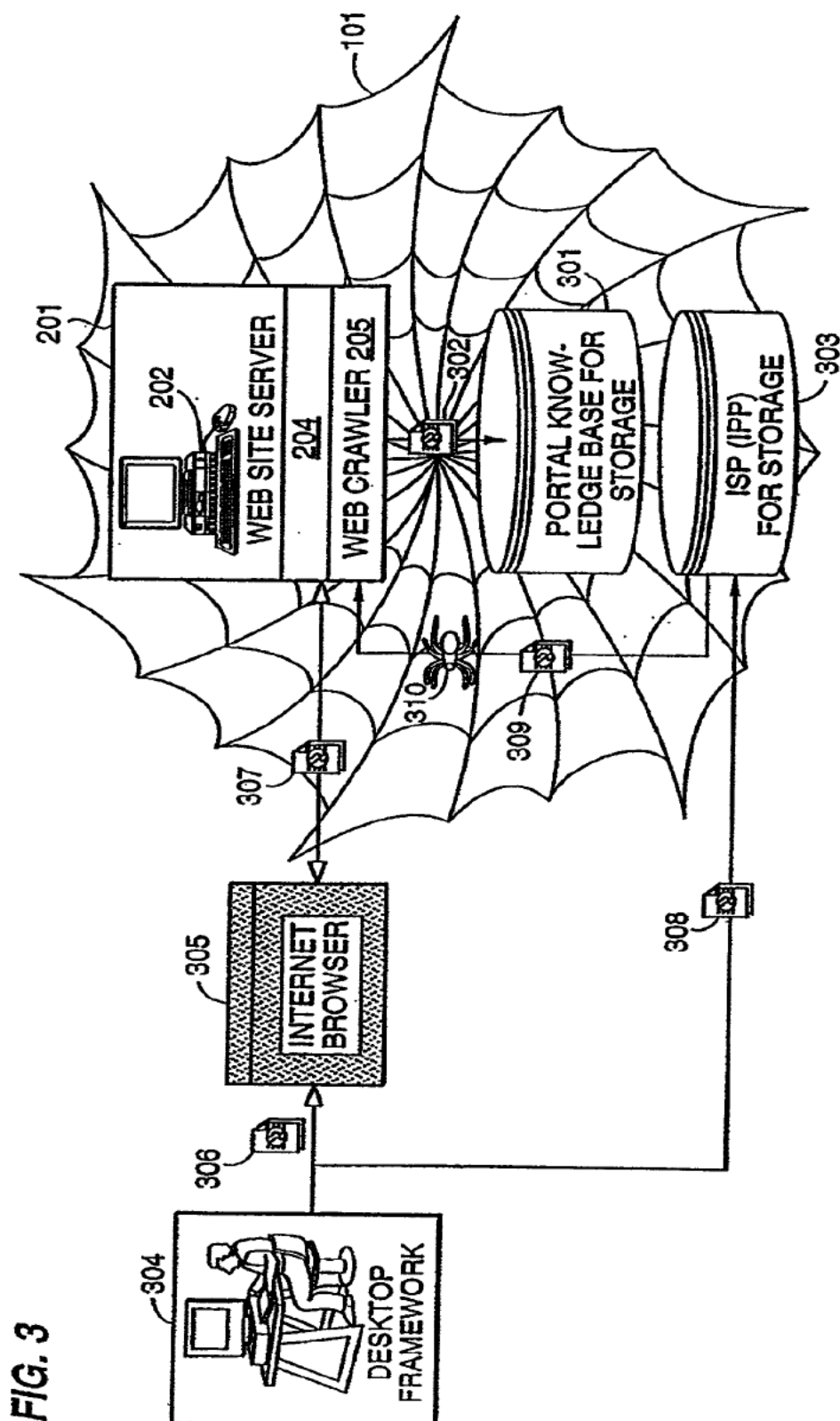
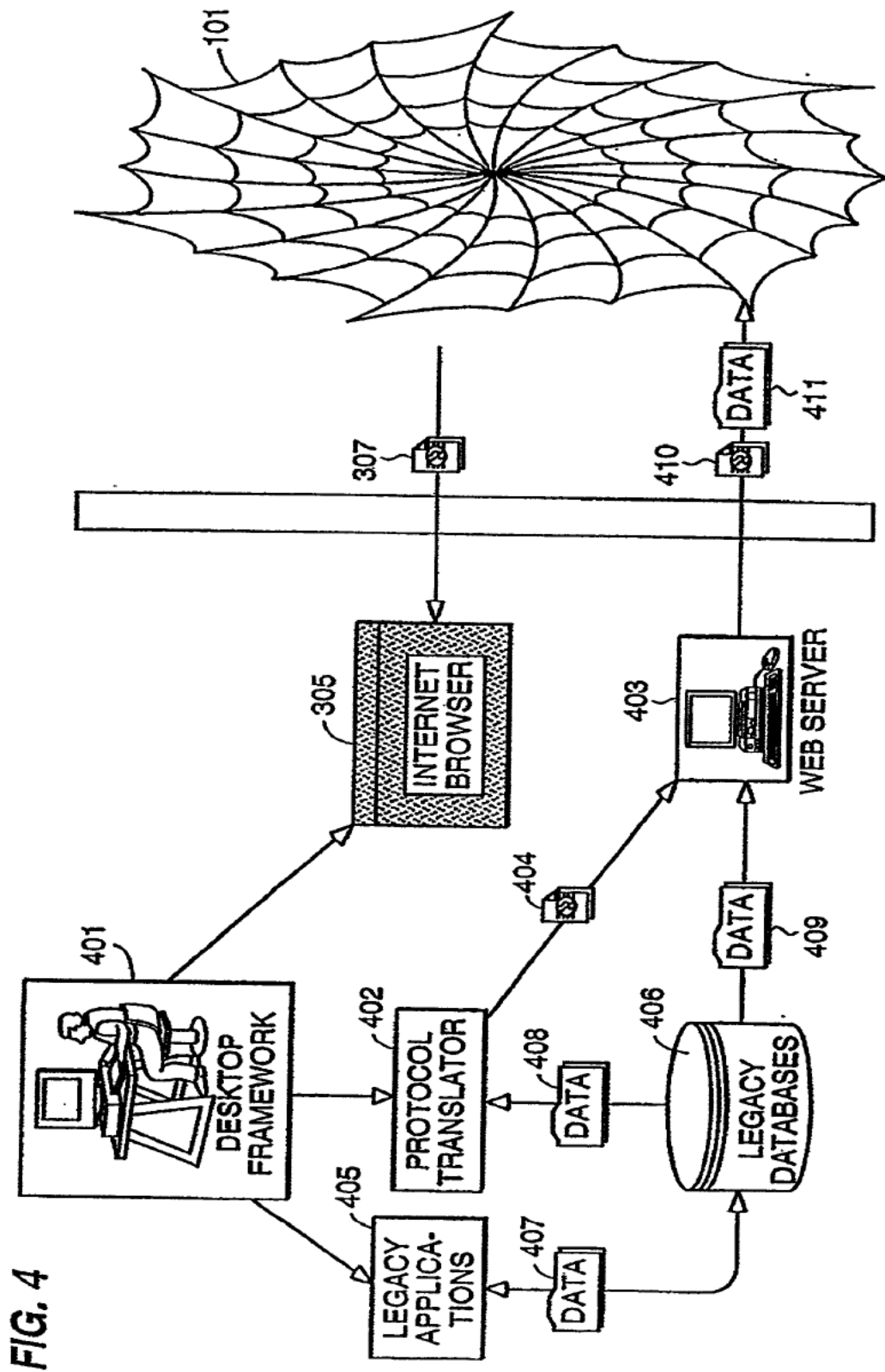
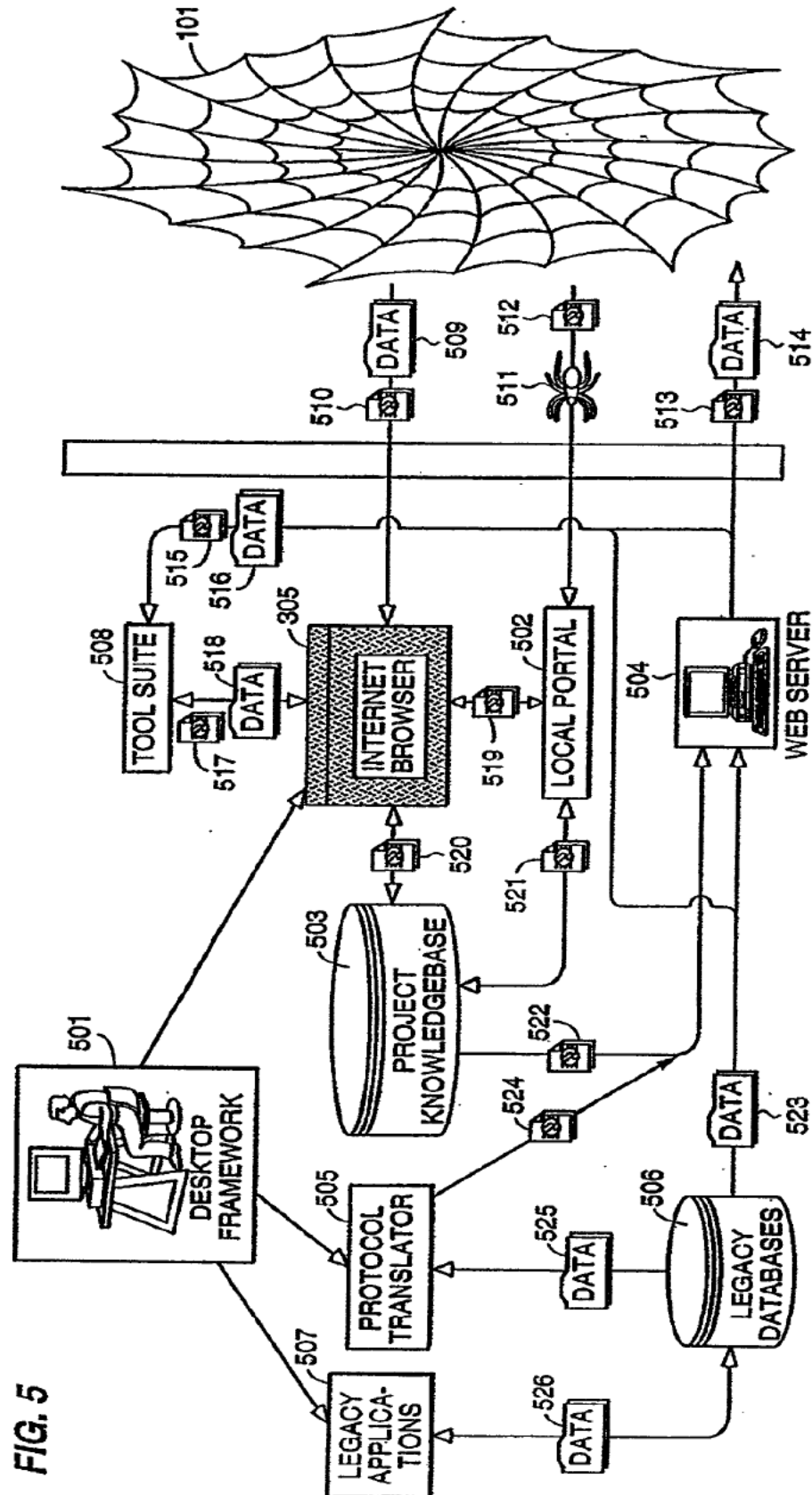
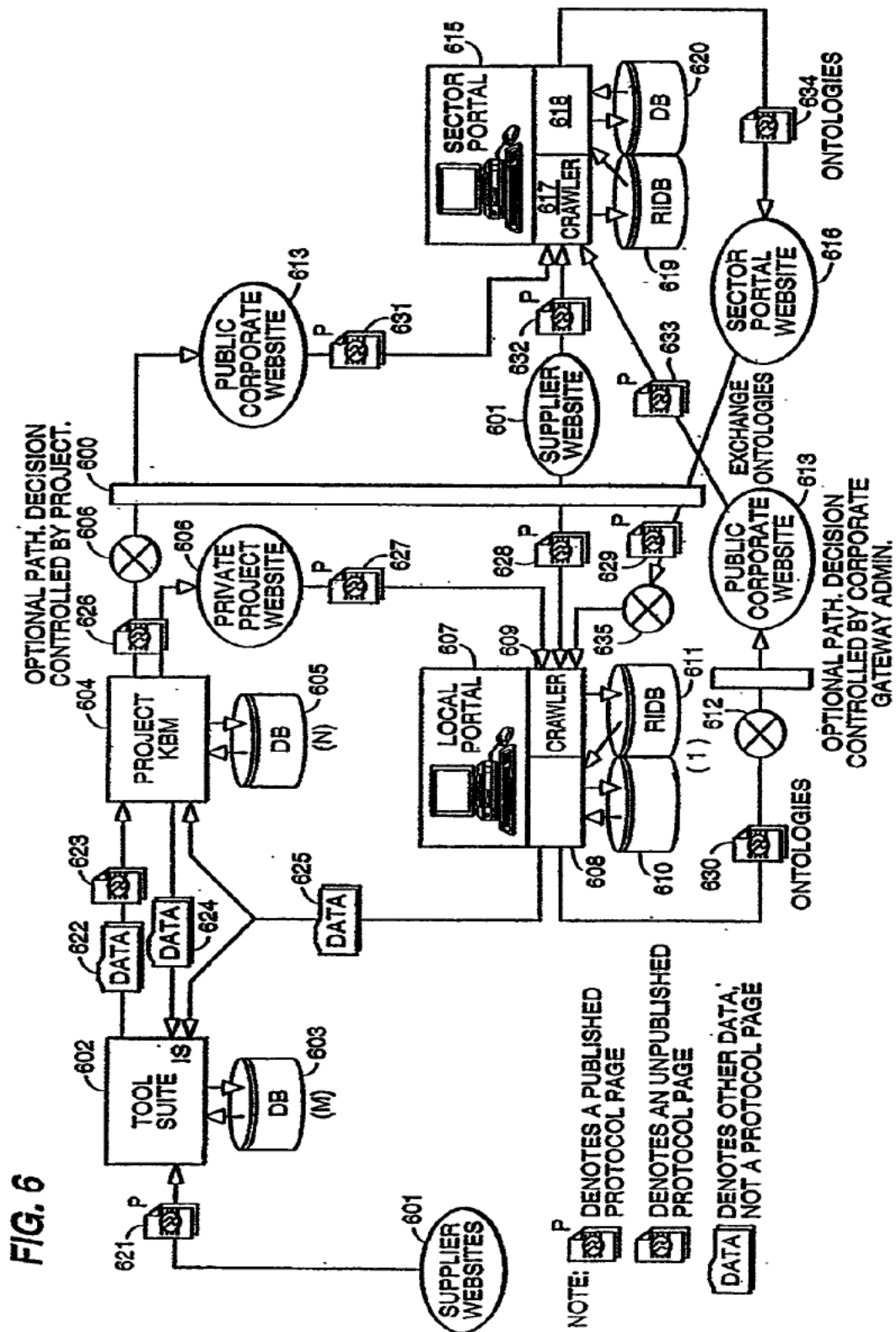


**FIG. 3**







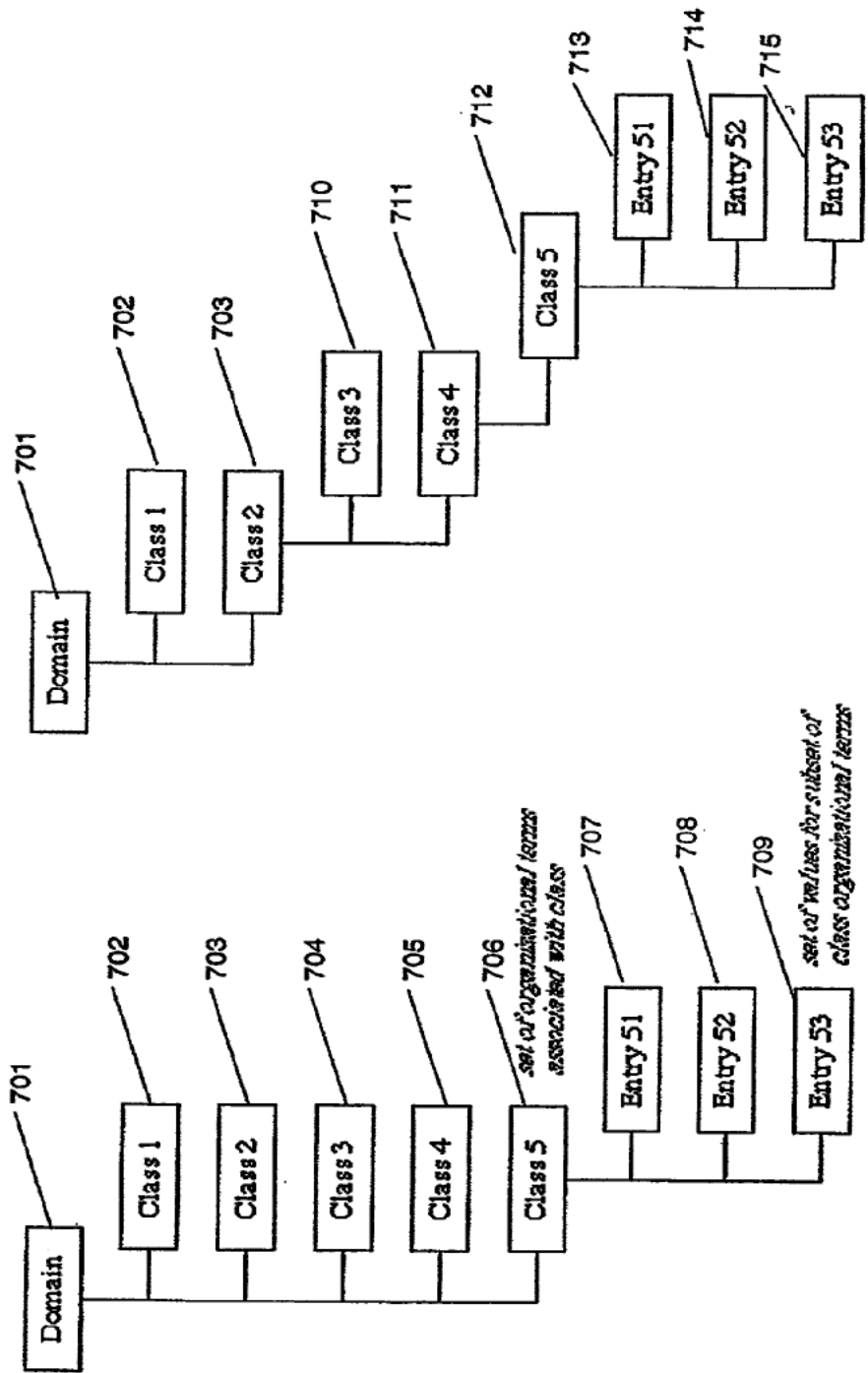


Figure 7A

Figure 7B

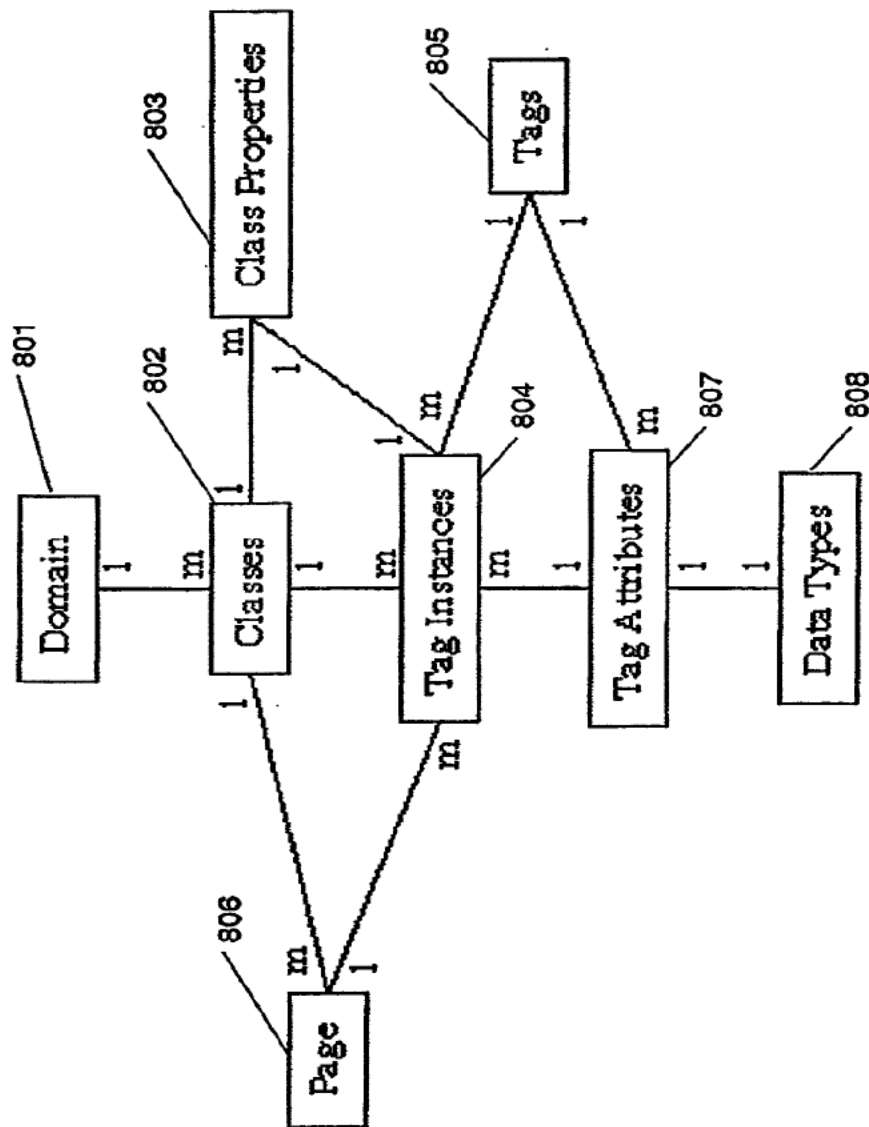


Figure 8

# SYSTEM, METHOD, AND MEDIUM FOR RETRIEVING, ORGANIZING, AND UTILIZING NETWORKED DATA USING DATABASES

## RELATED APPLICATION INFORMATION

[0001] This application is a continuation-in-part of U.S. Pat. Ser. No. 09/441,646, filed Nov. 23, 1999, which is a continuation of 09/120,182, filed Jul. 22, 1998, now U.S. Pat. No. 6,038,668, which is a continuation of U.S. Ser. No. 08/925,337, filed Sep. 8, 1997, now abandoned.

## BACKGROUND OF THE INVENTION

[0002] The following description includes some copyrighted material. While Applicants do not object to the copying of this specification for patent related purposes, Applicants reserve all copyrights to themselves and/or the assignee of the present invention.

### [0003] 1. Technical Overview

[0004] The present invention relates to a networked, information gathering and delivery system. More specifically, the present invention relates to a network based (intranet-based, Internet-based, extranet-based) cataloging system that provides a user with a plurality of methods of gathering and analyzing information relating to a subject of interest.

### [0005] 2. Related Art

[0006] Manufacturing and service industries exist to provide products and services to end users while making a profit at the same time. In this day and age, many businesses do not control the entire manufacturing process from collecting raw materials to outputting a high level end product. Rather, most businesses take existing components and incorporate the components into larger products. An example includes engineering design systems in which engineers take existing components and design new products using these components. To reduce costs, these businesses generally attempt to maximize profits through finding the most affordable materials and processes that satisfy their needs. Simultaneously, to meet the performance or quality objectives for the end product, business must find materials and processes that satisfy a variety of other criteria. Unfortunately, depending on the industry, finding the lowest cost materials or services which meet specified criteria can become very costly and time consuming.

[0007] Currently, the designers selecting the various components for a final design can quickly become overwhelmed by the available data as many lower end suppliers may produce similar products. Sorting the available products and attempting to match potential products into an overall design scheme quickly becomes a formidable task when a single end product may contain tens to hundreds to thousands of individual components with each component having its own set of associated data (for example, operating tolerances, power consumption, size constraints, etc.). Further, the manufacturing process, which produces the final design, may affect the end product in ways unconsidered by the product development team. Because of the significant burden placed on designers to handle design while another set of engineers generally handle the realization of the design through a manufacturing process, the end product may suffer from poor integration of the design and manufacturing processes. To this end, product support also suffers as the

team handling product support generally has minimal information regarding the actual design and manufacturing processes and how these processes affect the end product. At least one cause of the failures of the current design, manufacturing, and product support model is the requirement that all steps are handled by an engineer with minimal to no automated support for accommodating the demands of the current model. In fact, most information has to be manually integrated and considered in the formation of a final product.

[0008] Another failure of the current model is the lack of integrated tools available to a designer. Currently, a designer must wade through technical information and sequentially use disparate, unintegrated electronic tools to perform a variety of tasks such as requirements definitions, computer aided design (CAD), structural analysis, costing, manufacturing planning, etc. Accordingly, a need exists to automate and integrate the design function as much as possible.

[0009] The problems associated with the current design and manufacturing model exist in other industry sectors as well. For example, the real estate industry uses as its primary search and indexing tool the multiple listing service (MLS). This service allows real estate brokers to list properties as well as search for properties using a simple Boolean search function. However, the real estate agent is then required to peruse through the list of retrieved entries until a match is found to many diverse buyer requirements. Further, there is no ability to link to other sources of other information (for example, financial institutions providing loan information, etc.) that may be relevant to the information contained within the MLS. Moreover, MLS is limited to real estate agents only. Independent sellers and buyers cannot list and retrieve the listings available in MLS.

[0010] Unlike the restricted nature of MLS, the Internet provides easy access to a variety of information. Search engines (such as Lycos™, AltaVista™, and other search engines) exist that repeatedly scan the Internet (a.k.a., the World Wide Web or, simply, the Web) for content. Conventional search engines retrieve and store the textual content of HTML pages of the Web in large indexes. A user may later pass simple Boolean queries to databases created by the search engines and retrieve HTML pages relating to the submitted queries. The problem with such Boolean queries for these large indexes of stored information is that actually getting the desired information is cumbersome. Also, current systems must treat the found content as flat, unstructured data; for there is no way to create or find associations between the elements stored on HTML pages. Further, the appearance of subject-oriented searches fails to satisfy the needs of designers and like individuals interested in a narrow field of industry. To this end, current search systems are not directed to the needs of specific users.

[0011] A metadata framework has been proposed by Sakata et al. in *Metadata Mediation: Representation and Protocol*. Computer Networks and the Systems 29 (1997). Pp. 1137-1146. Sakata et al. discloses an ontology in a metadata framework. However, all navigation of the metadata appears to direct an end user back onto the original sources of the data on the Internet to receive requested information. Difficulties with constantly pursuing information located on multiple sites over the Internet include the occasional problems of the inability to contact desired sites due to communication breakdowns or bottlenecks. Accord-

ingly, while Sakata et al. provides a framework for creating and expanding a metadata protocol, it relies on continuous real-time access/availability to the distributed information located across the Internet for the underlying information for the end users and is, thus, open to delays and related problems.

#### SUMMARY OF THE INVENTION

**[0012]** The present invention relates to a computer-based information organizing, searching, retrieving, and exchanging platform. Through gathering organized information, a consumer accesses the gathered information to create an end product in a shorter period of time. The present invention uses a predefined common language and format (for example, hypertext markup language (HTML) extended to implement an underlying organizational structure or XML and related schema) for organizing information placed on the network of computers. When retrieved the organized information is readily sorted and cataloged into a repository of similar cataloged information.

**[0013]** When placing the information in a form readily accessible by other computers (for example, on the Internet), a supplier provides the information in a format including tags which characterize the formatted information. The format may include HTML, XML, and other mark-up languages. These tags may include a "class" identifier with each class having various "attributes" and "methods". The various attributes are contemplated to include "features" (comprising character strings), "parameters" (comprising numerical strings), and "index" tags. Through listing an item as belonging to a class and having various attributes, a supplier may fully identify the item using this structured language. Also, each class may be a nested item of another class. So, each component may be broken down into its elemental parts through classifying each of its parts individually. Notably, through the combination of various parts, products may be compiled. A similar approach can be applied to processes or services and their sub components. At this point, the consumer who compiled the product may place the product on the Web as specified in a page with the appropriate tags so that a later consumer may choose to use the previous consumer's product. In other aspects, the product information may be provided to a consumer using XML packaged with envelopes such as ICE (Information Content Exchange), WSDL (Web Services Descriptor Language), WAP and other information formats.

**[0014]** To this end, each consumer may also be called an integrator as each consumer may integrate parts into products for later integrators.

**[0015]** Also, some classes may have methods or tags associated with them as well which provide procedures for invoking simple or complex algorithms or computations and compute performance or behavior of an item in the class. Because the behavior of an item typically depends on its specific use and environment, these computations require information provided by the integrator and not knowable in advance by the product supplier. Thus, method tags provide access to procedures invoked by the integrator when he evaluates possible uses of the item. Method tags contain pointers to computational algorithms, and to the "signature" of the method. A signature defines what input and output information the algorithm needs and provides, respectively.

**[0016]** A scanning engine scans the computers having accessible pages to locate all pages having the predefined organizational structure as including class, attribute, and methods identifiers. Through organizing the retrieved information on the basis of the class, attribute, and methods information, the retrieved information is cataloged for later retrieval. The information may conform with various protocols (including but not limited to HTML and/or XML tags) that permit easier cataloging. Catalogs are collections of the retrieved information as stored in at least one computer system referred to as a "portal." The catalogs are stored and updated as needed. Instead of storing all possible information from each page or set of information from the remote information sources (as including graphic files, text files, Java scripts, and the like) or transmitted collection of information, the catalog, for example, may only store the class, attribute, and method identifier information that allow a user to search for and identify the accessed item. The pages or sets of information may include HTML web pages and other sets of information (including information marked up with the XML or related-markup language protocol). The remainder of the accessible information is left back on the originating site for later retrieval. Through this organizational scheme, high level descriptive information (for example, information that may satisfy an initial search) may be stored in a computer that interfaces between an end user and the information source. In this instance, the more detailed information may be left back at the originating source. Thus, the most detailed information, which is the most likely to change is readily updated at the source of the information, minimizing the information which has to be provided to keep each end user, as well as each interfacing computer, current.

**[0017]** As described above, in one embodiment, the cataloged information is stored in at least one local computer system interfacing between the end user and the information source. For purposes of simplicity, the local computer system as storing categorized information is hereinafter referred to as a "portal" as the portal acts as a primary interface to the organized information residing on the network of computers. When a portal is located on a user's local site, the user has quick access to the variety of information stored at the site.

**[0018]** To this end, embodiments of the present invention contemplate the sector portal maintaining a list of pages or content at each supplier's site that comply with a predefined protocol (HTML, XML, or the like) so that each portal need only retrieve a new list of supplier sites for compliant information, rather than search the Web as a whole for compliant information. All references herein to the protocol and to the placement of the protocol on pages are intended to be inclusive in that the protocol may include, XML, HTML, and other mark-up languages that provide information about the content of the transmitted information.

**[0019]** Also, for simplicity, the network of computers is referred to as the Web or the Internet. However, it is understood that any networked group of computers storing organized information is included within the scope of the invention. Further, while Web pages are referenced as one form of providing information to users, other techniques of providing information to users are specifically included. For example, one may use a device that does not display HTML pages (for example, some Internet-ready telephones as are

known in the art) but may display text resulting from a web-clipping service. Also, devices may display received information having been marked up using XML (extensible markup language).

[0020] The various users of the present invention may be categorized as information suppliers and information consumers. As will be described in greater detail below, as the information consumers in turn supply information to other upstream information consumers, every information consumer may be considered a supplier to other consumers.

[0021] To enable each supplier to provide requisite information on its pages, a "sector" portal establishes common terms (class, attribute, and method names) for the suppliers and consumers to use. The sector portal is so named because each industry sector is contemplated to have at least one governing portal from which all other portals in that industry sector derive their common vocabulary, taxonomy, or ontology (collectively, referred to herein as organizational terms). For reference, an ontology (as used herein) relates to a set of terms that are specified in some way to establish a conceptualization of a concept, specifying the names of the objects and characteristics of the concepts and indicating the hierarchical relationships of its pieces as is known in the art. Each sector portal has the ability to update its own list of accepted classes, attributes, and methods (generally referred to as descriptors). So, when a supplier begins to use new classes, methods, and attributes, the sector portals selectively update their ontology to include the new classes, methods, and attributes according to a predetermined scheme. For example, a predetermined scheme may include updating the ontology only after a new class, attribute, or method is used a predetermined number of times (e.g., ten times). An advantage of not updating the ontology for every possible new class, attribute, or method is that 1) mistakes will be made by suppliers in not using the predefined ontology so that the ontology should not necessarily include the mistakes of suppliers and 2) an ontology expanding too quickly may fail to minimize the number of terms used to describe a single item. To this end, the ontology should not necessarily include multiple listings of the same class or attribute when a single one may be used across suppliers.

[0022] Alternative definitions of an ontology are known and may be used in various implementations.

[0023] Once the catalogs are established at a portal, a consumer may use at least one of three search strategies to retrieve the stored information including Boolean searches, hierarchical searches (for example, where a consumer navigates through a hierarchy of classes), and parametric searches (which may include fuzzy searches).

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0024] FIG. 1 shows an organization diagram of various suppliers and consumers according to embodiments of the present invention.

[0025] FIG. 2 shows the configuration of an industry sector portal according to embodiments of the present invention.

[0026] FIG. 3 shows a low-end supplier with browser access according to embodiments of the present invention.

[0027] FIG. 4 shows a high-end supplier with a page generator according to embodiments of the present invention.

[0028] FIG. 5 shows a consumer according to embodiments of the present invention.

[0029] FIG. 6 shows an architectural representation of sector and local portals, according to embodiments of the present invention.

[0030] FIGS. 7A-B show relationships between items in an ontology according to embodiments of the present invention.

[0031] FIG. 8 shows relationships in a database model according to embodiments of the present invention.

#### DETAILED DESCRIPTION

[0032] The present invention will now be described in detail with reference to the accompanying drawings. While the present invention is described in the context of a communications network including a specific number of components, the cataloging search and retrieval system and integrated tool suite according to the present invention may be incorporated into networks of many structures and sizes. The drawings are intended to provide one example of a network configuration in which a cataloging search and retrieval system may be implemented and are not intended to limit the applicability of the present invention to other network configurations.

[0033] FIG. 1 shows a high level functional arrangement of the contributors and users of the organized information. The network of computers, in this example, is shown as the Internet or Web 101. Portal 102, in one aspect of the present invention, functions as an entry way into the stored catalogs of information. As will be described in greater detail below, portal 102 may reside at a number of locations. Also multiple portals may exist for an industry. In the case of a single portal 102 for an industry, the portal 102 is referred to as the sector portal 102. In the case of multiple portals 102 (for example, each company having its own portal 102), then one of the portals 102 may be considered the sector portal. The sector portal may maintain control over the accepted ontology. Portal 102 is contemplated to include at least one processor and at least one storage device. Through the implementation of stored program modules, the portal (or portals) 102 may control the collection and dissemination of information as described in greater detail with respect to FIGS. 2-6.

[0034] Also shown in FIG. 1 are suppliers (104 and 105) of information. The suppliers (104 and 105) of information provide information to portal 102 for use by user 103. User 103 uses the information stored in portal 102 to make decisions on the information as well as retrieve greater amounts of information, as needed from suppliers 104 and 105. Further, user 103 includes a tool suite which uses the information in portal 102 and at providers 104 and 105 to provide additional functions to the user including, inter alia, the availability of alternative decision trades, the likelihood of success of selected components meeting an overall design criteria, etc. To provide information to portal 102, suppliers 104 and 105 make their information available for retrieval via the Web 101 by making available such information through Web pages in HTML, XML, or other markup language format or other Web-accessible storage media using similar formats on a server connected to the Web 101. As only some of the entities (business, organizations, etc.)

which access the Web 101 have the ability to provide the server technology to Web-enable their information, suppliers 104 and 105 are separated into two groups: suppliers (104) who do not have the capability to provide Internet access and the suppliers (105) who do have the capability. As for suppliers (104) who cannot Web-enable their information, the system allows suppliers 104 to upload their product information (and other information) to an ISP for page hosting purposes. Accordingly, suppliers 104 are considered low-end suppliers. The functionality of uploading pages from suppliers 104 is treated in greater detail with respect to FIG. 3. As for suppliers (105) who can Web-enable their information, the system accommodates their ability as well. The functionality of Web-enabling from suppliers 105 is treated in greater detail with respect to FIG. 4.

[0035] FIG. 1 further shows user 103 using information stored in portal 102. Embodiments of the present invention contemplate portal 102 not storing every piece of available information in portal 102 as the amount of storage space required may quickly become overwhelming. So, portal 102 stores some information that may answer some initial questions for user 103 and point to suppliers 104 and 105 for additional information. As user 103 consumes the information provided over Web 101, user 103 may be referred to as an integrator as it integrates the available information. Notably, the resulting data sets combined by user 103 may be placed in another portal 102 for another user's access. In this instance, user 103 becomes another supplier. For simplicity, the groupings of the entities using the disclosed system are referred to collectively as "players". The labels applied to the various players in the system (supplier, user, consumer, integrator, etc.) are intended to be relative to the surrounding players and not intended to be an absolute definition of the player.

[0036] To provide information to portal 102, suppliers 104 and 105 (and user 103, when applicable) encode their pages using a predefined protocol. Use of the protocol encourages placing available information in an organized format. The protocol may include tag codes which describe the information contained therein. For example, an HTML-based tag protocol may include the following information:

---

```
<UC domain="product" class="gimbals/shanks/widgets">
<UC_ENTRY name="standard widget" realization state="available">
<UC_FEAT name="Supplier Name" value="Acme Corporation">
<UC_FEAT name="Manufacturer Name" value="Acme Corporation">
<UC_ID name="part number" value="1234">
<UC_PAR name="Price" value="$100">
<UC_PAR name="Length" value="2.5" units="inches">
<UC_PAR name="Width" value="3.5" units="inches">
```

---

[0037] The tag <UC\*> (where "\*" may include additional tags) is the identifier to portal 102 that the following page is in an organizational format useable by portal 102. It is readily apparent that any predefined tag could be used as long as it is recognizable by the portal. For simplicity, organized language used by suppliers and read by portal 102 will hereinafter be referred to as "the protocol". The protocol may include various fields that describe both the information as well as the supplier of the information. In the above example, the "domain", "class", "entry name", "realization",

and "PAR" (parameter) fields relate to the information supplied. The FEAT (feature) fields relate to the supplier of the information. In the above example, the Acme Corporation is both a manufacturer and supplier of a standard widget (a product) (currently available) which costs \$100 per unit and has dimensions 2.5 inches by 3.5 inches. The domain field relates to the type of product. In general, the feature field may relate to alphanumeric character strings while the parameter field may relate to numerical values.

[0038] While HTML-based information encodings are known, HTML-based tags may have restrictions including the following:

[0039] 1) Information needs to be encoded twice on the same page: once for the protocol data content (the protocol tags ignored by an HTML browser) and once for the browser display (HTML tags). Modifying page content thus requires a tool that can concurrently recognize and edit corresponding HTML and protocol tags.

[0040] 2) By mingling content and display, a page containing content needs to be modified if the display is to be modified, even if the content remains the same. The corresponding process can be unwieldy, can introduce unintended corruption of protocol data, and can lead to inconsistencies between the protocol data and its independently defined display.

[0041] 3) The inventory and storage requirements of static pages can become large, with considerable duplicate information both within and across the body of static pages.

[0042] Alternatively, one may use XML and/or stylesheets to provide a more unified approach to handling information. For example, the information may solely contain XML content as shown in the following code segment:

---

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="/UCLP.xsl"?>
<UC domain="Products" version="2.0" class="Fuze/Proximity/Stand
Alone/RF/Guided Missile" status="current">
<UC_ENTRY name="6665005" realization state="available">
<UC_ID Name="Product Name" value="Target Sensor"/>
<UC_FEAT name="QualifiedTo" value="ISO9001"/>
<UC_ID name="Supplier Number" value="94990"/>
<UC_PAR name="Price" value="Proprietary" units="Dollars"/>
<UC_ENTRY>
</UC>
```

---

[0043] Using XML, a system receiving information may rely on display instructions that are maintained separately in stylesheets, wherein the XML content being refer to and being displayed using these stylesheets. The XML tags may be stored as static "pages" or the tags may be generated dynamically through use of protocol translators that map between the portal ontology and the native format of the data store that contain the content to be identified as protocol data. All references herein to the protocol and to the placement of the protocol on pages are intended to be inclusive in that the protocol may include, XML, HTML, and other mark-up languages that provide information about the content of the transmitted information.

[0044] There are a number of advantages for using XML over HTML, including:

[0045] (1) The need for duplication of protocol data on a static HTML page is removed;

[0046] (2) Inconsistency between protocol data and its display is removed;

[0047] (3) The volume of page inventory is significantly reduced; and,

[0048] (4) The XML design of the protocol enables the dynamic generation of protocol data.

[0049] Embodiments of the present invention contemplate including a variety of different domains, each directed to at least one type of industry sector. For explanatory purposes, a variety of industry sectors are contemplated and, to these sectors, overlapping domains may exist. For example, the manufacturing industry sector may include products as well as processes. The shipping industry may include timetables for shipping routes (overlapping as to delivery schedules for the manufacturing industry). The real estate industry may include commercial and residential properties. The loan industry may include commercial and residential loans as well as the risk factors associated therewith.

[0050] The class field relates to an ordered hierarchy of information. In this example, the standard widget falls into subclass "widget", which is a subclass of subclass "shanks", which is a subclass of class "gimbals". In an alternative example, a real estate property listing may belong to a subclass "locality", which is a subclass of class "residential properties".

[0051] FIG. 2 shows a generic portal 201. Generic portal 201 is represented in the middle of Web 101 as the portals may be located in a variety of different physical and functional locations, depending on the industry and the resources available to the players involved. Portal 201 includes a Web site server 202, an online page generator 204 (also known as a protocol translator), a searching system 205 (referred to herein as a Web Crawler), search engines 203, and a knowledge base 206. Here, knowledge base 206 is represented as an industry sector knowledge base. Embodiments of the present invention contemplate at least one industry sector knowledge base existing for each industry in a given locality. For example, in the real estate industry, a single knowledge base may support the entire industry in given locality, as real estate agents would access a single knowledge base for retrieving residential and commercial listings. In other industries, each company may include its own updateable knowledge base allowing limited access to the public and greater access to employees.

[0052] Web crawler 205 periodically scans the Web for pages as represented by crawler 210. Pages 209 are retrieved and parsed in Web crawler 205. As will be described in greater detail below, the parsed pages 208 containing a usable organizational structure (for example, identified as including the <UC> tags) are stored in knowledge base 206 for indexing and future retrieval. The knowledge base 206 stores pages 208 based on the classes found therein as well as the attributes and methods found therein (for example, the length, width, and other properties of the contained infor-

mation). When a specific page needs to be retrieved for information contained therein, search engines 203 search knowledge base 206 for results meeting various search criteria. As knowledge base 206 stores previously parsed pages 208 in an ordered manner (via class, attribute, and method information) a variety of different searches may be performed on the stored information.

[0053] To search for information, embodiments of the present invention contemplate a user submitting a search request via an HTML page 211 forwarded to server 202. The server then forwards the requested information to search engines 203. A variety of different searches may be performed.

[0054] For example, a hierarchical search may be performed. Using a hierarchical search engine, a user may select a general class and step through the various subclasses until a specific subclass is reached. For instance, a user may be looking for brushless motors as classified in Parts/electromechanical parts/motors/brushless motors (where the general class is "Parts" and "electromechanical parts", "motors", and "brushless motors" are nested subclasses). To find a list of brushless motors, a user would first select "parts", then "electromechanical parts", then "motors", and then "brushless motors". When reaching "brushless motors", the list of available brushless motors would be displayed. Also, a full text Boolean search may be performed on the terms stored from Web Crawler 205. For example, using a search engine accessing a linear index (for example, as supplied by Alta Vista™ from Digital Equipment Corporation), a user may request information on "brushless motor". In response the search request, the search engine would scan through a stored index of terms for "brushless" and "motor" and retrieve all matching results.

[0055] Finally, a user may request a parametric search. A parametric search is a search that attempts to match input parameters with known values in an arithmetic sense. For example, a parametric search may deal with numerical values and character strings, but generally only through a Boolean comparison. Fuzziness allows matches to be made based on the "closeness" of one value compared to another or one character string compared to another. To be able to compare character strings, a fuzzy logic search relies on a predefined value map, which defines what closeness means for the attributes in question. By referencing the value map a parametric search can handle both numerical and non-numerical searches.

[0056] In a parametric fuzzy logic search, a user supplies target values for the attributes of interest (e.g., the selected parameters, features or indices), together with the relative importance of each selected attribute and other ranking information (such as "more is better" or "closer is better"). The search engine would then return the data objects that best match the target values. An example of a fuzzy logic search engine includes a search engine developed by the Inso Corporation of Massachusetts. One may use, for example, a qualitative statement or statements that describe closeness of one object to another. Examples include ISO 9000 different levels of quality, geographical locations converted to distance or converted with constraints (for example, on same side of river), and the like. An example of ranking various fields is shown below:

| Subclass:         | Rectifiers     |                |                |       |
|-------------------|----------------|----------------|----------------|-------|
| Rating:           | IF AV          | VRRM           | IFSM           | TJ    |
| Desired Values:   | 5              | 75             | 215            | 180   |
| Relative          | very           | little         | very           | most  |
| Importance:       |                |                |                |       |
| Ranking Criteria: | more is better | more is better | less is better | exact |

[0057] Through implementing a fuzzy search specifying the above values and qualifiers, the system attempts to find the desired rectifier with the most important values satisfied first.

[0058] The web site server 202 converts the received results from search engine 203 into HTML and forwards or "serves" the created page to the requesting user through page 211.

[0059] To assist in the retrieval of information from portal 201, a user may access a variety of different tools. As described above, the user may access the search engines 203 through submitted page 211. Also, the user may employ the use of data agents located on the user's terminal (or computer) to pre-format the submitted information to include, modify, or remove submitted information as necessary. Data agents may include CGI scripts, JAVA components, ActiveX components, and the like. In addition to data agents residing on server 202, data agents may also reside with the suppliers 104 and 105 to provide enhanced processing of requested information for a user.

[0060] FIG. 2 also shows protocol translator 204 as part of the portal 201. The translator 204 facilitates supplier publication of HTML pages that are compliant with the protocol and the industry common vocabulary or ontology. This is described further in the discussion of FIG. 3.

[0061] FIG. 3 shows the low-end supplier 104 of FIG. 1 as low-end supplier 304 providing information to be scanned by a portal 201. Portal 201 is equivalent to that described with respect to FIG. 2. FIG. 3 also includes an ISP 303 (an Internet Service Provider) in Web 101. Low-end supplier 304 (a supplier who lacks the capability or desire to support organized pages locally) retrieves pages 307 into an Internet browser (software running on user 304's computer which translates HTML pages and displays the pages on user 304's display).

[0062] Pages 307 include at least one template for populating and submitting back onto Web 101. Embodiments of the present invention contemplate template pages 307 including at least one of an initial set of class, attribute, and method identifiers for population. For example, user 304 requests from portal 201 the design template for electric motors. In reply, protocol translator 204 retrieves the desired class/subclass, attribute, and method ontologies. These ontologies are translated into an HTML form and sent to user 304's Internet Browser 305 as template pages 307. User 304 next populates the received template pages 307 as completed template pages 306.

[0063] Pages 306 may be forwarded back directly to portal 201 where the pages 306 are parsed and added to knowledge base 301 as page 302. Alternatively, if user 304 has an ISP 303 that will publish the populated template pages 306, then

user 304 forwards the populated template pages 306 as formatted web pages 308 to ISP 303 for storage. Next, web crawler 310 scans ISP for complying pages 308 and retrieves them as pages 309 for storage in knowledge base 301 for indexing and later retrieval.

[0064] Pages 307 may be sent from a variety of different portals 201. For example, in an industry sector where only one portal exists, the template page 307 would be sent from the single portal 201. In industry sectors where multiple portals exist, each supplier and/or user may wish others to adapt a specific variety of ontology that suits their business. However, conflicting ontologies may result. To alleviate problems created by varying ontologies across a single industry sector, different portals may be programmed to respond to a specific variety of protocols. To this end, embodiments of the present invention may contemplate a single sector portal governing the recognized ontology. While each of the other portals are free to adapt their ontology as need be, retrieving template pages 307 from a sector portal almost guarantees compliance with the other portals that may have varied from the accepted ontology. Further, if a non-sector portal 201 (for example, a private corporation's portal) receives a request for a template, embodiments of the present invention contemplate providing the user 304 with a choice of sector-wide ontology or the private corporation's ontology. Moreover, variations of the type of ontology provided exist depending on specifics of user 304 including specifics such as whether the user 304 is cleared to receive certain classified (i.e., secret) ontologies (where the classes, attributes, and/or methods themselves are classified), etc.

[0065] Once portal 201 has received new pages, a variety of methods exist for updating the other portals in the sector (provided that the industry supports other portals). First, each portal 201 may include a list of all other portals. Once new information is added, the portal 201 may transmit a message to all other portals in the industry that new information has been received by portal 201 and either attaching the new information or providing a location on portal 201 for access to the stored information. This method provides the advantages that a user may transmit information to a minimum number of portals and have the transmitted information replicated by the receiving portal. Second, the sector portal may receive and maintain a listing of all new pages. Periodically, or upon demand of a user, each portal may scan the sector portal for new page addresses then scan the new pages themselves for retrieval of the desired class, attribute, and method information.

[0066] The information which a supplier may have on his site may include image data including VRML data, CGM Active Graphics images, Java Parametric Optimizers, CAD drawings, performance specifications, executable simulations, links to the supplier's online ordering system, links to previous order status, or other image data information, etc. Also, non-image data may be found at the supplier's site including HTML pages including organized protocol, EDI links (links furthering electronic data interchange), technical data, technical services, or other non-image data information, etc. When a portal scans a supplier's site, the invention contemplates that at least some technical data (and possibly some image data) may be retrieved and stored in the portal. The remaining technical and image data may remain stored at the supplier's site for later access. The supplier may also

use data agents to gather information from the users in order to, among other things, enhance the usability of the supplier's site for the user. These enhancements may include personalizing the supplier's site through placing data agents on the supplier's site including possibly CGI scripts, Java applets, and interfaces to the supplier's databases. For example, a supplier's site may include a Java applet that may enable enhanced searching of a supplier's site for green widgets. By providing the Java applet to a scanning portal, the portal may incorporate the Java applet into its stored index of suppliers so that, when information is requested from the supplier's site, the user may run the Java component to enhance the retrieval of information (for example, through the user dynamically accessing and manipulating a supplier's information as stored in the portal). Also, the Java component may format information when transmitted from the user to the supplier so, when received at the supplier's server, the supplier may quickly process received questions as having been previously formatted by the Java applet.

[0067] FIG. 4 shows a supplier 105 that has the capability to publish its own protocol-compliant pages. The supplier 105's site includes a desktop framework 401, an Internet browser 305, protocol translator 402, web server 403, legacy applications 405, and legacy databases 406. An operator at desktop framework 401 requests a template page 307. The template page may be transferred to the protocol translator where the template page is combined with data (class, method, attribute, etc.) specifying the supplied products and processes from supplier 105. The resulting page 404 may be transferred to web server 403 for posting. On request, the compiled pages 404 may be transferred to other sites (for example, users' portals 102) as pages 410. Alternately, the protocol translator may map data stored in legacy databases 406 to fields in the template, create an HTML page 404 embodying that data and the special protocol tags, and transfer that page 404 to web server 403 for posting. The translator may also save the mapping for repeated use.

[0068] Legacy databases 406 may store additional data (for example, images, VRML representations, specifications, or other data, etc.) that describe in greater detail supplier 105's products and/or processes. Supplier 105 may make this additional information available to other portals as required (for portals meeting various requirements, possibly including security or other requirements). For example, when further information not found or in addition to information found on pages 410, the web server receiving the request may forward the request to legacy database 406 from where the requested information is accessed. The legacy database 406 may forward the data 409 to web server 403 which then may forward the requested data, properly formatted in HTML as data 411. Additionally, the data 411 may be transmitted in a form compatible with the applet at the requesting portal 102 for proper display in conjunction with the requesting applet.

[0069] Other applications 405 may provide support for other functions of supplier 105's site including updating of the information stored in legacy database 406 as well as for allowing supplier 105's site to support a suite of integration tools as necessary for applications contained in tool suite 602 (described in greater detail below with respect to FIG. 6).

[0070] FIG. 5 shows user site 103 including a desktop framework 501, an Internet browser 305, a local portal 502,

a web server 504, project knowledge base 503, protocol translator 505, legacy applications 507, legacy databases 506 and tool suite 508. Internet browser 305 may function similarly to the browser of FIGS. 3 and 4 yet may retrieve data 509 in addition to pages 510. Browser 305 may transmit the received pages to tool suite 508, project knowledge base 503 and local portal 502. Also, browser 305 may exchange the received data 509 as needed with tool suite 508.

[0071] The desktop framework 501 may access a variety of integration tools, data agents, PDM's, the local portal 502, an application programming interface (API), and an internal cache of accessed pages.

[0072] Local portal 502 may retrieve pages 512 through crawler 511 which may scan the Internet 101. The retrieved page may then be parsed to page 521 and transferred to project knowledge base 503. When accessed, project knowledge base 503 provides requested information as page 520 to Internet browser 305 (when, for example, an operator requests information stored in project knowledge base 503) or as page 522 to web server 504 (when, for example, page 522 containing a finished product or process is to be published onto the Internet 101, for other upstream portals 102). When web server 504 serves to publish compiled pages, it may act like server 403 in FIG. 4. To this end, legacy applications 507, protocol translator 505, legacy database 506 may function similarly to that of the corresponding elements in FIG. 4.

[0073] Tool suite 508 may also receive additional pages 515 from web server 504 or pages 517 from browser 305. Data 516 and data 518 may follow similar supply paths respectively. Tool suite 508 includes a variety of tools which integrate the information contained in other portals 102, the local portal 502, and web server 504 (as retrieving information, at least in part, from legacy database 506).

[0074] FIG. 6 shows a more detailed illustration of the system of FIG. 5. Delineation 600 indicates the boundary between a corporate portal and associated systems and entities on the Internet. Local portal 607 includes at least two distinct systems: a crawler 609 (similar to crawler 205 described above in FIG. 2), and ontology tracking component 608. Connected to the crawler 609 and ontology tracking component 608 are a reverse index database (RIDB) 611 and a database 610. Database 610 may be an object oriented database (OODB) or a relational database (RDB). Crawler 609 may receive pages from a number of sources including a supplier's web site 601, a sector portal web site 616 (publishing pages and/or ontologies 634), or a private project web site 606. Crawler 609 parses the received pages into character strings and stores the strings in RIDB 611 where RIDB 611 indicates from where the strings were received. When RIDB 611 has received a predetermined amount of protocol data, the protocol data is sent to ontology tracking component 608 where the ontology tracking component may store the found protocol data in database 610. In storing the found protocol data in database 610, ontology tracking component 608 may control the database 610 to store the received protocol data in an organized structure. For example, the ontology tracking component 608 may control the database 610 to store the received protocol data grouped by class, attribute, and/or methods. Also, the database 610 may store the signatures separate from the methods to allow indexing. When requested, ontology tracking com-

ponent **608** may search and retrieve information from database **610** to satisfy search requests as outlined above.

[0075] One feature of the ontology tracking component **608** is that it may maintain the ontology for organizing database **610**. When new classes, attributes, and/or methods are received by crawler **609**, the new information is provided to ontology tracking component **608** for processing. Ontology tracking component **608** may process the new class/method/attribute in a variety of ways. For example, ontology tracking component **608** may not update its own ontology, instead waiting for standardized ontology **634** from sector portal **615**. In this regard, a single ontology is maintained with sector portal **615**. When all portals subscribe to only allowing sector portal **615** to update the accepted ontology, an invariant ontology is maintained. Alternately, ontology tracking component **608** may always update its own ontology, allowing maximum flexibility and creativity to the authors of published pages. Further, ontology tracking component **608** may selectively update the ontology based on a predetermined pattern of usage or time associated with newly received classes/methods/attributes. For example, ontology tracking component **608** may wait until a new class has been used in at least 10 pages before it dynamically updates itself to allow others to use the new class as well as index on the class. Also, ontology tracking component **608** may wait until a new class has been used for 3 months before updating its own ontology. By limiting the rate at which the ontology tracking component **608** updates itself, the ontology tracking component **608** may minimize spurious occurrences of new terms (which, in some instances, may be due to typographical or transmission errors). Once updated, ontology tracking component may provide the new ontology to sector portal **615** through public corporate web site **613** via page **633**.

[0076] In addition to the ontology described above, the ontology tracking component **608** may also include the following baseline for associated objects:

|              |                                                    |
|--------------|----------------------------------------------------|
| Ensemble     | A product which includes this product;             |
| Ancestor     | Previous version of this product;                  |
| Descendent   | Updated version of this product;                   |
| Component    | A component used in this product;                  |
| Accessories  | Other products which go the with the product; and, |
| Alternatives | Other products which might also meet requirements. |

[0077] Baseline ontology methods may include business functions, price or cost analysis, calculation of lead time required for product delivery, calculation of need for customization, a function for ordering parts, and a function for ordering services.

[0078] Additional methods may be performed by portal **607** including page certification, a crawler report of visited pages, and a page validation to ensure errors are kept to a minimum.

[0079] Integration with other system components such as the tool suite or project knowledge base may be made through CORBA or JAVA as needed.

[0080] The information stored in database **610** may be considered atomic or elemental in nature that it is combined with other information stored in the OODB (other atoms or

elements, collectively "items") and provided to a user. The user requests the items from portal **607** and receives data **625**. Project knowledge base manager **604** receives data **625** and stores the data in database **605**. Here, the received data **625** may be cataloged according to current projects on which a user is working. Also, the database **605** may also contain catalogs of items. For example, a user may want to build a device that requires a first component and a second component. The user searches via a search query (Boolean, hierarchical, parametric, etc.) and receives items A and B that match the search criteria for the first component and receives items C and D that match the search criteria for the second component. The project knowledge base manager may store the information relating to items A and B together in a first catalog in database **605** and items C and D together in a second catalog in database **605**. Accordingly, when the user needs to evaluate or try a different item (part, process, etc.), the user may directly access the stored list of catalogs. Also the current project may be stored in database **605** as a work in progress. The difference between the information stored in a catalog and the information stored as a work in progress is that the catalogs store end items, while information stored, as a work in progress is a combination of elements or end items. At least one goal of the catalog is to provide a readily accessible database of information for use in the creation of end items, while at least one goal of the project database is to store the created end items (products or processes) as they evolve.

[0081] Once a user completes an end item, project knowledge base manager **604** may direct information relating to the completed end item (completed page **626**) to private project web site **606** (which may contain additional end items (products, processes, etc.)) for internal distribution to local portal **607**. As local portal **607** may be one of a plurality of local portals **607** for a corporation, the information provided on private project web site **606** may be transmitted to these additional web sites as well. One embodiment of the present invention contemplates local portal **607** scanning private project web site **606** periodically or upon request from a user. Periodic scanning keeps local portal **607** updated autonomously. On the other hand, scanning upon request minimizes bandwidth consumption for exchanging information. To this end, a user may specify when the portal is to scan (the web in general or specific sites) for new protocol-compliant pages.

[0082] An example of a process implemented in portal **607** is as follows:

[0083] 1) the portal **607** scans for information (**628**, **629**) from external information providers (**601**, **616**) and for information (**627**) from internal information providers (**606**);

[0084] 2) the portal stores the received information;

[0085] 3) the portal dynamically creates web pages for a user based on the stored information; and,

[0086] 4) provides links to the additional information left back on the supplier's page.

[0087] Tool suite **602** includes various tools which may use information **624** provided by project KBM **604**, information **625** from the ontology tracking component **608** and related databases **610** and **611**, and from supplier web sites **601**. These tools may include a requirements integration and

verification tool (which assures that assembled design objects (end items) meet individual requirements and comply with predefined rules); an affordability monitor (which determines if the cost of the included items and the processes required for combination exceeds predetermined budgets), a trade-off manager (which trades various elements to arrive at varying end items so to assist in the decision making process), a realization generator (which creates process alternatives for combining elements for determining whether the end item may be improved through alternative processes), and an integrated supply chain scheduler (which coordinates the communication and information exchange among entities in a supply chain to arrive at feasible manufacturing schedules). These tools may autonomously access the information stored in the portal 607 or stored in supplier web sites 601 to compile and assist users in defining end items.

[0088] The database of FIG. 6 may be constructed with an object-oriented database. In an alternative embodiment, the database 610 may be replaced by a relational database. Some observations have been made with respect to OODBs in comparison to relational databases. Within an OODB, a class object is created and defined to comprise a set of organizational terms. The organizational terms in turn are defined using protocol tags.

[0089] Each class instance prescribes tag values to define the set of organizational terms corresponding to that class. Protocol data then describe the entries that are instances of each class by prescribing values for some subset of organizational terms defining the class. At the highest level, classes can be grouped into domains of interests. This is illustrated in FIG. 7A for a single domain. FIG. 7A shows a domain 701 with classes 1-5 (702-706) directly depending on or descending from domain 701. All classes have the same rank. Class 5 further includes entries 51-53 (707-709).

[0090] The structure in FIG. 7A is one example of a structure that may be used by the object-oriented system (including database 610). FIG. 7B is another example of a data structure that may be used. In FIG. 7B, Class 2703, Class 3710, Class 4711, and Class 5712 are not independent but are connected through some hierarchical structure. For both structures shown in FIG. 7A and FIG. 7B, entries 51-53 (707-709 and 713-715) may be assigned to any class node 702-706 and 711-713. For simplicity of illustration only, entries are assigned only to Class 5.

[0091] While the two structures of FIGS. 7A and 7B may be adapted to provide information, FIG. 7B provides the additional benefit of having classes within classes. This provides for a richer set of relationships to be stored within the domain and, therefore, impart more information to a user when retrieved. For example, using FIG. 7B, one may store a "hex bolt" class, which is dependent from a "bolt" class, which is dependent on a "fastener" class.

[0092] While the structures of FIGS. 7A and 7B may both be stored in database 610, the data structure of FIG. 7A lends itself to easier searching for an object oriented version of database 610. For instance, a class identifier that is defined in terms of a text string, which is the intuitive name supplied by a human user, may be searched. Searching for a particular class based on such a text string may be impractical in the multi-tier hierarchy of FIG. 7B because each hierarchy branch may have to be searched in turn. For

example, the search engine traversing the content of an object oriented database searches the class identifier for class 1702, then searches the entries subordinate to class 1702 to see if those entries (which may be other classes) satisfy the search string. This repetitive searching each level of entries off each class increases the number of items that need to be searched. The flat hierarchy of FIG. 7A provides an advantage over the structure of FIG. 7B in that all classes are at the same level. So, searching the data structure is easier in that only one level of classes needs to be searched.

[0093] To permit faster searching for an object oriented database traversing the data structure of FIG. 7B, the data structure of FIG. 7A may be combined with a separate but complementary hierarchy that may be provided to capture class relationships such as those shown in FIG. 7B. Alternately, indexes such as those commonly used in a relational database management system (referred to as a RDBMS) may also maintain such relationships. Routine conversion of an OODB to an RDBMS may result in difficulty because a relational design classically requires a static data structure and the services of a RDBMS administrator to modify the structure as the class hierarchy evolves.

[0094] Database 610 maybe implemented using an RDBMS, with the following advantages:

[0095] (1) enables the object class structure defined for the OODB version of database 610;

[0096] (2) supports flexible ontology methods defined above without requiring the services of a RDBMS administrator; and

[0097] (3) combines the ontology defined above and the corresponding hierarchy into one unified structure.

[0098] By use of this approach, the definition of an ontology may comprise (1) a hierarchical organization of domain content and (2) properties associated with each hierarchy class. The property information is conveyed through use of the protocol, where the protocol comprises tag definitions through which organizational terms may be captured, where such organizational terms describe characteristics/descriptors of a class. Relative position of the class within the hierarchy can be one of the organizational items.

[0099] A high level relational database model including key relationships among tables is shown in FIG. 8. Each of the boxes in FIG. 8 represents a separate table in the database. Entries in the Tag Attribute table 807 include the attributes that are available to define protocol tags. Each Tag Attribute entry corresponds to one Data Type (from table 808).

[0100] The Tags 805 are defined as having 1 . . . m (one-to-many) relationship to Tag Attributes 807. Protocol tags are defined in terms of the of the relationships between entries in the Tags and Tag Attributes tables.

[0101] A domain is defined as a top-level entity in the Domain table 801. Each entry in the Domain table 801 corresponds to one domain. A domain may comprise 1 . . . m classes (reflected in class table 802).

[0102] Entries in the Class Properties table 803 comprise the properties that are available to define classes. Classes are defined in the classes table 802 as having 1 . . . m Class

Properties from class properties table **803**. Classes are defined in terms of the relationships between the entries in the Classes and Class Properties tables (**802** and **803**, respectively).

**[0103]** A page **806** is a set of protocol data, expressed through tags defined in the Tags table **805**, and representing values for class properties corresponding to the class for which the page is providing protocol data. An entry in the Page table **806** comprises 1 . . . m Tag Instances **804** and each page contains tag instances for only one class. 1 . . . m pages comprise the total set of protocol data corresponding to a class. Each tag on a page **806** is an entry in the Tag Instance table **804**. There are 1 . . . m tag instances per page and each tag instance corresponds to one entry in the Tags table **805**. Many tag instances across multiple classes can use any tag. Class properties may be entered manually into the Class Properties table **803**, the class properties may be accumulated from new entries in the Tag Instances table **804**, or a combination of both methods can be used.

**[0104]** The following is an example of the use of the tables of **FIG. 8** for the search of class identifiers. A relational database is used for database **610**. A search request is received by the portal **607** and processed. The relational database is searched by first querying domain table **801** for the proper domain. From the received result, classes table **802** is queried for all classes in the appropriate domain. The first class referenced by the class table **802** is located and its class identifier searched. The search then proceeds through the entries in the table and searches the class identifier of each referenced class. The other tables **803-808** may be accessed as well during searches.

**[0105]** The use of the RDBMS of **FIG. 8** provides the following advantages:

**[0106]** (1) The RDBMS design encompasses major benefits of the object-oriented structure attributed to OODB version of database **610** as described above and of relational indices which support capturing and searching the ontology hierarchy;

**[0107]** (2) The RDBMS design has expanded market potential as companies generally have relational databases on site, rather than object-oriented databases;

**[0108]** (3) The RDBMS design supports greater scalability as the number of classes, corresponding class properties, and complexity of the hierarchy grow;

**[0109]** (4) The expanded ontology of the RDBMS design enables a more consistent underlying architecture by unifying the previously defined ontology and the corresponding, but previously separate, hierarchy.

**[0110]** Using the above described invention, users may rapidly configure a supply chain of information, easily search the chain for relevant information, determine the best match of existing end items to new requirements, integrate desired goals or preferences into a created end item, aggregate product realization plans for assembling a product, estimate the feasibility and effect of information trade-offs as pertaining to, for example, costs, facilitate the exchange of models, simulations, and statistical information, integrate and manage trades, and exchange information with other applications.

**[0111]** Applications of the above described invention may include manufactured goods relating to, at least, industry sectors including aerospace, automotive, general and discrete manufacturing, electronics, pharmaceuticals, and other industry specific manufacturing supply and design chains and to service industries including, at least, real estate markets, financial institutions' information exchange, loan institutions' information exchange, insurance institutions' information exchange, and other industry specific information exchanges.

**[0112]** While the present invention has been particularly described with reference to the preferred embodiments, it should be readily apparent to those of ordinary skill in the art that changes and modifications in form and details may be made without departing from the spirit and scope of the invention. For example, the processes described above may be implemented in software. For instance, the portals' **102** ability to retrieve, organize and store may be coordinated from stored memory modules as executed by said portal. It is intended that the appended claims include such changes and modifications.

We claim:

1. A portal for retrieving information from a supplier, said supplier providing information arranged in a predefined protocol to said portal, said portal comprising:

a storage for storing said information from said supplier in catalogs, said storage comprising a database, said information being stored in said database in accordance with organizational terms defined by the protocol; and

a processor for querying said storage in response to searches from users.

2. The portal according to claim 1, wherein said database is a relational database.

3. The portal according to claim 1, wherein said database is an object oriented database.

4. The portal according to claim 1, wherein the protocol defines a domain and a class, and

said relational database includes a domain table and a class table.

5. A system for accessing catalogs as stored by a portal in a database, said system comprising:

an output for generating a search request and transmitting said search request to said portal;

an input for receiving the results of a search of said database,

wherein the portal performs said search of information stored in said catalogs as stored in said portal and wherein said catalogs are created using organizational terms from suppliers.

6. The system according to claim 5, wherein said database is an object oriented database.

7. The system according to claim 5, wherein said database is a relational database.

8. The system according to claim 7, wherein said relational database includes a hierarchical structure.

9. A method for retrieving information from a portal, said portal having a database having a hierarchical structure of classes and entries in which information from suppliers is

stored based on organizational terms present in the information from said suppliers, said method comprising the steps of:

transmitting a search request to said portal; and  
receiving a search result from said portal.

**10.** The method according to claim 9, wherein said database is a relational database.

**11.** The method according to claim 9, wherein said database is an object oriented database.

\* \* \* \* \*