PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 17/30

A1

(11) International Publication Number: WO 99/57659

(43) International Publication Date: 11 November 1999 (11.11.99)

(21) International Application Number: PCT/US99/09741

(22) International Filing Date: 4 May 1999 (04.05.99)

(30) Priority Data:
09/073,895 6 May 1998 (06.05.98) US

(71) Applicant: DATAFUSION, INC. [US/US]; 139 Townsend Street, San Francisco, CA 94107 (US).

(72) Inventors: SCHATZ, Joel; 601 Buena Vista West #4, San Francisco, CA 94117 (US). BUTLER, Mark; 808 Gooding Drive, Albany, CA 94706 (US). EDHOLM, Kaj; Apartment X, 811 West California Avenue, Sunnyvale, CA 94086 (US). FREELAND, Mark; 1556 Sanchez Street, San Francisco, CA 94131 (US). NORGARD, Barbara; 230 12th Avenue #304, San Francisco, CA 94118 (US). SCHOLER, Stephan; 1870 Washington Street #10, San Francisco, CA 94109 (US). SMITH, Fraser; 1740 Bush Street, San Francisco, CA 94109 (US).

(74) Agents: CALDWELL, Gregory, D. et al.; Blakely, Sokoloff, Taylor & Zafman, 7th floor, 12400 Wilshire Boulevard, Los Angeles, CA 90025–1026 (US).

(81) Designated States: AE, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

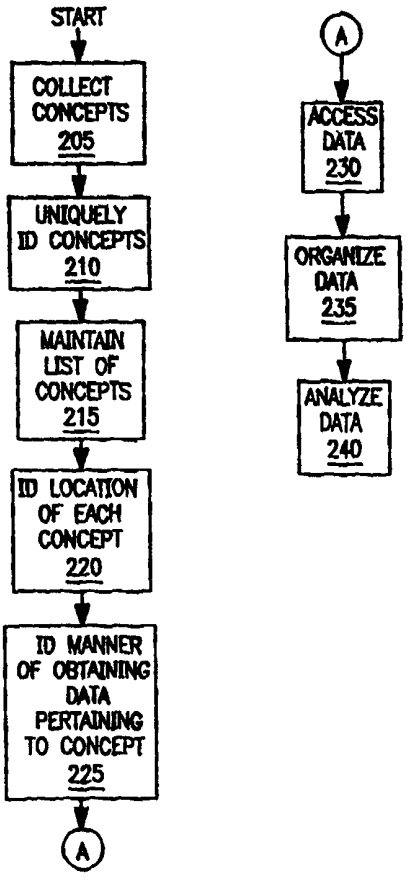
With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD AND APPARATUS FOR COLLECTING, ORGANIZING AND ANALYZING DATA

(57) Abstract

A method and apparatus for collecting concepts (205), compiling metadata indicating the location and manner of accessing data relating to the concepts (220), accessing the data at the location in accordance with the manner for obtaining the data as indicated by the metadata (225), organizing and interconnecting the concepts in a knowledge module (235), and performing analysis on the data pertaining to the concepts so organized in the knowledge module (240).



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHOD AND APPARATUS FOR COLLECTING, ORGANIZING AND ANALYZING DATA

COPYRIGHT NOTICE

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention is related to a computer implemented method for collecting, organizing and analyzing data, as may be used, for example, in conjunction with a modeling system. In particular, the present invention relates to a method for collecting concepts, compiling metadata indicating the location and manner of accessing data relating to the concepts, accessing the data at the location and in accordance with the manner for obtaining the data as indicated by the metadata, and performing analysis on the data.

Description of the Related Art

Prior art tools are available for building knowledge modules, handling metadata, or maintaining meta-thesauri. Prior art tools also exist for data retrieval or computer modeling. Indeed, sophisticated implementations of such tools are commercially available. However, the prior art presently lacks integrated tools that combine knowledge modules, metadata, and meta-thesauri

into one system. In particular, prior art systems directed to knowledge mapping lack the ability to consistently organize underlying concepts using a metathesaurus that also points to underlying information sources via metadata. Conversely, prior art systems that are directed to maintaining meta-thesauri or metadata lack the ability to allow the user to represent domain knowledge via an interactive knowledge module. What is needed is an integrated method and apparatus for the identifying, accessing, collecting, organizing, and analyzing information, in particular, by integrating knowledge modules with a metathesaurus that contains concepts used to label elements in the module as well as to associate data sources in the metadata with module elements. Further, the metadata would facilitate the accessing of information and mechanisms in the knowledge module to facilitate the analyzing of information and computer modeling.

BRIEF SUMMARY OF THE INVENTION

The present invention provides a method by which concepts are collected, accessed, organized and analyzed. The concepts and relationships between the same may be utilized to create a knowledge module (KM) comprising entities and links for analysis. A KM author selects concepts from a concept catalog, organizes the concepts into a knowledge module, accesses data pertaining to the concepts, and performs the analysis. Data exchanged between KM entities are selected and represented as links between the entities. The links are organized in such a manner so as to identify conceptual relationships between entities. The entities and links may be depicted in a network diagram. An embodiment of the present invention further provides for a method for interconnecting two or more knowledge modules to form a single merged knowledge module.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

The present invention is illustrated by way of example and not limitation in the following figures. Like references indicate similar elements, in which:

Figure 1 is a block diagram as may be used by an embodiment of the present invention.

Figure 2 is a flow chart of an embodiment of a method contemplated by the present invention.

Figure 3 is a block diagram of an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention generally relates to a method and apparatus for collecting concepts, compiling metadata indicating the location and manner of accessing data relating to the concepts, accessing the data at the location and in accordance with the manner for obtaining the data as indicated by the metadata, organizing and interconnecting the concepts in a knowledge module (KM), and performing analysis on the data pertaining to the concepts so organized in the KM. In the following description, numerous specific details are set forth in order to provide a thorough understanding of the present invention. However, it will be apparent to one of ordinary skill in the art that the present invention may be practiced without these specific details. In other instances, well-known structures, architectures, and techniques have not been shown to avoid unnecessarily obscuring the present invention.

Definitions

A review of the following terms is useful for an understanding of the present invention.

Authoring: the process of manipulating Metacode™ to build or edit knowledge modules (KMs).

Concept: a thing, process, idea, or notion that may be cataloged.

Concept Catalog (CC): a repository for records that describe, define and label concepts in a natural language. Each record in the CC comprises a universally unique identification number (concept ID, or CID), a textual label, and links to source classifications from which the concept is derived. Concept IDs from the CC are utilized in knowledge module (KM) building in order to enable KM interconnection and association of KM components with data sources. CIDs are further used in a metadata database (MDDB) as an index to relate concepts to sources of information or data about those concepts. The CC may be supported by multiple lexicons, including variant spellings, abbreviations, acronyms, etc.

The present invention contemplates a master concept catalog that contains concepts used in the public domain, hereinafter also referred to as "the concept catalog." Such a master CC is necessary for CIDs and concept labels to be consistent in all contexts. Other CCs may provide concepts having different CIDs. However, KMs containing CIDs from a CC other than the master CC may not be able to interconnect with other KMs containing CIDs from the master CC because the CIDs are likely to be inconsistent.

Knowledge Module (KM): a manifestation of some aspect of a KM author's knowledge constructed as a set or network of entities and links encoded in accordance with an embodiment of the present invention. A KM contains at least one entity. KMs are fungible, in that a KM author can extract a portion of one KM and treat it as another, or connect two KMs to form a

third KM. Entities and links are each instances of concepts. Entities in a KM may contain Mechanisms.

Knowledge Module Authoring Tool (KMAT): a computer software program tool used by KM authors to construct a knowledge module, providing a user interface for the KM author, for example, windows-based dialog boxes and graphics. A KM author may use the KMAT to select concepts from the concept catalog (CC) and attach them to entities, or links. In accordance with an embodiment of the present invention, the KM author may browse the CC or the metadata database (MDDB) and attach a concept or a metadata database entry (i.e., data pointer) to a KM component. The KMAT may be used to create a KM from scratch, edit an existing KM in the knowledge module catalog (KMC), or insert/combine a KM into another KM.

Knowledge Module Catalog (KMC): a repository of knowledge modules (KMs) available to KM authors through the KMAT. A KM author may edit an existing KM in the KMC either replacing the existing KM or creating a new one,, or insert/combine a KM or KM components into another KM in the KMC. The KMC is also the repository of KMs available to users for viewing.

Knowledge Module Components: entities or links in a Knowledge Module.

Knowledge Module (KM) Entity: an instance of a concept that stands alone in a knowledge module. KM entities are different from concept labels in the concept catalog. Entities include a label and mechanism box.

Knowledge Module (KM) Interconnection: the connection of two KMs by adding links between the entities in the KMs, or by establishing hierarchical relations between entities from

the two KMs. KM interconnection is facilitated by comparing the CIDs attached to entities and links in each KM.

Knowledge Module (KM) Link: an instance of a concept which forms a connection between two KM entities. A link may be either a data link or a semantic link. The former represents the transferring of data in or out of mechanisms inside an entity, the latter qualifies the relationship between two entities. KM links may include a label and a concept.

Mechanism: rule or procedure for transforming data. An entity in a KM may comprise a mechanism box (or simply, mechanism). A mechanism box comprises a set of functions from input values and internal variables to output values or internal variables. Retrieved or external data can enter the computation in the mechanism box through association with internal variables.

Metacode™: a protocol for associating concept IDs (CIDs), metadata, and data with knowledge module (KM) components, transferring data to underlying analytical models, and facilitating KM interconnection.

Metadata Database (MDDB): a database comprising information about sources of information or data. The MDDB indicates an information source that relates to an entity or link in a knowledge module, as well as where to locate it, how to access it, its format, the manner in which it is maintained, the assumptions underlying it, an indication of its reliability, the cost associated with using it, and so on. For information sources that can be queried (e.g., databases), the MDDB acts as a gateway to retrieving data for associating with, or inserting into, a knowledge module. The MDDB may be a centralized or distributed database, and each of its records comprises one or more concept IDs (CIDs) corresponding to the concept catalog (CC).

Taxonomy: a set of rules for classification of objects or things. As a manifestation of an ontology, a taxonomy is a particular view of how things should be grouped together or classified. Taxonomy is also the process of classification, including the methods used and assumptions made.

Hardware Overview

Referring to Figure 1, a computer system upon which an embodiment of the present invention can be implemented is shown as 100. System 100 comprises a bus or other communication means 101 for communicating information, and a processing means 102 coupled with bus 101 for processing information. Processing means 102 may be comprised of one or more processors. System 100 further comprises a random access memory (RAM) or other dynamic storage device 104 (referred to as main memory), organized as either shared memory or distributed memory if in a multiprocessor architecture, coupled to bus 101 for storing information and instructions to be executed by processor 102. Main memory 104 also may be used for storing temporary variables or other intermediate information during execution of instructions by processor 102. Computer system 100 also comprises a read only memory (ROM) and/or other static storage device 106 coupled to bus 101 for storing static information and instructions for processor 102. Data storage device 107 is coupled to bus 101 for storing information and instructions. A data storage device 107 such as a magnetic disk or optical disk and its corresponding disk drive can be coupled to computer system 100.

In a preferred embodiment of the invention, computer system 100 is configured with a network interface 103 for coupling computer system 100 to a data communications network such as a corporate intranet, the Internet or World Wide Web graphical portion of the Internet, to access information for

purposes of performing analysis on the information and the like. Browser software and the like is provided for querying and displaying to the user information obtained from intranet or Internet accessible databases.

Computer system 100 is coupled via bus 101 to a display device 121, such as a cathode ray tube (CRT), head mounted display, etc., for displaying information to a computer user. An alphanumeric input device 122, such as a keyboard including alphanumeric and other keys, is typically coupled to bus 101 for communicating information and command selections to processor 102. Another type of user input device is cursor control 123, such as a mouse, trackball, or cursor direction keys for communicating direction information and command selections to processor 102 and for controlling cursor movement on display 121. This input device typically has two degrees of freedom in two axes, a first axis (e.g., x) and a second axis (e.g., y), that allows the device to specify positions in a plane. Additionally, the input device 123 may have three degrees of movement such as a three-dimensional spaceball device and may be utilized to specify positions along, e.g., three axes such as an x, y and z axis.

Alternatively, other input devices such as a stylus or pen can be used to interact with the display. A displayed object on a computer screen can be selected by using a stylus or pen to touch the displayed object. The computer detects the selection by implementing a touch sensitive screen. Similarly, a light pen and a light sensitive screen can be used for selecting a displayed object. Such devices may thus detect selection position and the selection as a single operation instead of the "point and click," as in a system incorporating a mouse or trackball. Stylus and pen based input devices as well as touch and light sensitive screens are well known in the art. Such a system may also lack a keyboard such as 122 wherein the interface is provided

via the stylus as a writing instrument (like a pen) and the written text is interpreted using optical character recognition (OCR) techniques.

In the currently preferred embodiment of the invention, computer system 100 is configured to execute a database application. Computer system 100 may be one of many computer systems accessing data stored in the same database, which may be centralized or distributed. Each of the computer systems may be executing one or more transactions. The mechanisms of a database management system execute by using memory structures, permanent data storage structures, and processes. The memory structures exist in main memory 104 of computer system 100. It should be noted that in a distributed database management system, the memory structures may exist in the main memory of one or more computer systems that constitute the database management system. Processes are jobs or tasks performed by processors in response to executing sequences of instructions stored in the memory of the computer systems.

Overview of an Embodiment of the Present Invention

With reference to Figures 2 and 3 and the above defined terms, an overview of an embodiment 300 of the present invention is provided. The present invention controls computer system 100 to build a concept catalog. Building the concept catalog involves the steps of collecting concepts at step 205. Each of the concepts are associated with, preferably, a globally unique concept identifier (CID) at step 210. A data structure, i.e., concept catalog (CC) 315, is constructed at step 215, as a repository for the concepts so collected. In addition to associating a CID with each concept, each concept is further associated with a textual label. It should be noted

that while each concept in the CC is associated with a unique CID, the label may apply to more than one concept.

Having established a concept catalog, indexed by both textual labels and CIDs, the next steps involve identifying at least the location of each concept at step 220, and identifying the manner of obtaining data pertaining to each concept from the identified location, at step 225. Steps 220 and 225 are central to the creation of metadata as maintained, for example, in metadata database (MDDB) 320. A metadata database editor 330 builds the metadata database, in which each entry comprises a CID associated with one or more concepts from the concept catalog 315. The CID is used as an index to information about the concepts with which it is associated in the concept catalog. It should be noted that although the MDDB 320 is illustrated in Figure 3 as a single data structure, it is appreciated that the MDDB 320 may well be a distributed database. A data analyst operating the metadata database editor 330 to build MDDB 320 accesses data storages, e.g., databases 340, to obtain information regarding the location of data, and other relevant information regarding the data, to be maintained in MDDB 320.

Given a concept catalog through which to browse and from which to search and select concepts, and a metadata database providing the necessary information for querying and obtaining data pertaining to the selected concepts, the present invention facilitates the creation of a knowledge module 335 for analysis of the concepts and the relationships between them. At step 230, concepts are selected from the concept catalog by a knowledge module (KM) author, and data relating to the concepts may be accessed in accordance with the instructions and information regarding such access as indicated by the MDDB 320. Alternatively, since the MDDB maintains CIDs for the concepts that it provides information about, the author may

access the MDDB for both concepts and information about accessing the concepts. A knowledge module authoring tool (KMAT) 305 provides the author with the ability to build KM 335, using, for example, dialog boxes and graphics, extracting CIDs and/or labels from the CC and/or MDDB in establishing entities and links for the KM. Furthermore, the KMAT may also be used to attach data or metadata to entities and links to enhance the KM or convey a specific message.

The knowledge module catalog (KMC) 310 also may provide input to the KM 335 via the KMAT 305. The KMC provides a repository for KMs that the KM author may access for building new KMs. Additionally, the KMC provides a repository for storing KMs created by the KM author. The KM author can create new KMs and store them in the KMC, open and edit existing KMs, replace an existing KM or create a new one, or insert one or more KMs from the KMC into another KM. In any case, the KMs stored in the KMC may be used as building blocks for larger and/or more detailed knowledge modules.

At step 235, the KMAT 305 organizes the data obtained in step 230, for purposes of analyzing the data at step 240. It is appreciated that steps 230-240, although illustrated in the flow diagram of Figure 2 as a sequential series of steps, may be, and in fact most often are performed iteratively. KM 335 may comprise data relating to selected concepts, CIDs corresponding to the selected concepts and, optionally, MDDB records (pointers to data) relating to selected concepts. Concepts are manifested in the KM as either entities or links, the latter of which may be characterized as semantic links that qualify an entity to which they are attached, or data links that indicate the exchange of data between entities. In organizing the data in the KM, links are made between any two entities in accordance with a protocol, referred to herein as

Metacode™, and described in more detail below. After organizing the data, a viewer may analyze the KM via KM viewer 325. The viewer may optionally access either in-house or external data pertaining to the entities and links provided in the KM from storage media, e.g., databases 340. The viewer may optionally execute mechanisms as specified in the mechanism boxes of the KM entities.

Concept Catalog

As described above, the concept catalog (CC) 315 uniquely and unambiguously identifies things, ideas or notions referred to herein as concepts. The unique identification of concepts in the concept catalog provides for the retrieval of information relating to the concepts, and the interconnection of separate knowledge modules. Moreover, the CIDs and labels in the CC are provided and incorporated into the MDDB 320 to index information regarding the location of data pertaining to the associated concepts.

Concepts that may be incorporated in the CC 315 are derived from many different sources of information; the different sources of information likely organize information according to different classification schemes for distinguishing concepts. The concept catalog therefore provides translation allowing for concepts classified according to one scheme to be mapped to another classification scheme despite the use of different labels (e.g., corn farming versus corn production). An embodiment of the present invention contemplates the translation process incorporating lexical matching tools to facilitate the process of mapping labels to each other.

The concept catalog provides for both formal and informal classification schemes. Formal classification schemes generally provide definitions of nodes and syndetic structure, i.e.,

pointers to related concepts. Informal classification schemes, e.g., dictionaries, relational database structures, etc., generally provide definitions only. The concept catalog of the present invention is capable of accommodating both formal and informal classifications. Additionally, the concept catalog supports multiple viewpoints of the organization of concepts. For example, different classifications organize the same concepts in different manners (e.g., the parent of a node in one classification scheme can be the child of the same node in another classification scheme), or a node may focus on different attributes for the same concept (e.g., a chemical characterized in structural terms versus functional terms). Furthermore, the concept catalog does not itself provide any links between concepts even though it allows the mapping of concepts in one classification scheme to concepts in another.

A KM author searches the concept catalog when building a knowledge module for analysis. The concept catalog provides for the searching and retrieval of a specific concept by CID, and of synonyms and quasi-synonyms (i.e., concepts close in meaning but not exactly the same) relating to the specific concept and appropriately labeled. Given that the concept catalog is developed from multiple source classifications, it is structured such that a KM author can move from a concept or node in the concept catalog to the corresponding node in the source classification and display the node within the context of the syndetic structure of the source classification.

Metadata Database

The metadata database (MDDB) 320, as indicated above, describes data resources, their location, and the means of accessing the data resources. Thus, the MDDB is essentially a set of pointers to data. An embodiment of the present invention contemplates a MDDB that comprises

a list of publicly available and relevant information. As with the concept catalog, the MDDB grows over time as those who maintain the MDDB update it to point to newly found data. Additionally, unlike the concept catalog, it may shrink, as referenced data vanishes, and this is reflected in the MDDB. Alternatively, the MDDB utilizes data access filters that allow for automated retrieval of data stored in distributed, heterogeneous systems. The resources pointed to by the MDDB include but are not limited to computer accessible sources such as databases, spreadsheets, World Wide Web sites, and knowledge modules, which may contain data in various forms, including but not limited to video, sound, structured text, and plain text. Additionally, the MDDB may point to collections of information such as archives, libraries, file cabinets, books, and conference proceedings, etc. Each metadata record comprises information pertaining to a data source, such as what data is available, how to retrieve it, cost of retrieval, time for retrieval, quality of the data resources, whether and how the data resources may be manipulated, etc. Although the present invention contemplates searching the MDDB via the KMAT, it is appreciated that the MDDB may be searched on a stand-alone basis.

Knowledge Module Authoring Tool

The knowledge module authoring tool (KMAT) 305 provides the means by which a KM author may access the concept catalog and metadata database to build a knowledge module. In one embodiment, the tool operates on computer system 100 to provide a windows-based graphical user interface for receiving input from the KM author. The KMAT supports at least the common file operations associated with application software such as necessary to support the present invention in directing computer system 100 to create a knowledge module. Likewise, the KMAT supports editing capabilities such as select, copy, paste, insert and delete to manipulate

the organization and interconnection of entities and knowledge modules with semantic and/or data links in KM 335. Semantic and/or data links may be made, according to the Metacode ™ protocol, between any two entities in a KM. The protocol does not specify guidelines or rules regarding how the entities relate to each other. How the KM author views the relationship between entities depends on the purpose and intent of the KM and analysis performed thereon.

CLAIMS

What is claimed is:

1. A computer implemented method of analyzing information, comprising the steps of:

a) collecting a plurality of concepts;

b) uniquely identifying each of the plurality of concepts with a concept identifier;

c) maintaining a list of the collected plurality of concepts;

d) identifying a location of data pertaining to each of the plurality of concepts;

e) selecting concepts from the list of collected plurality of concepts;

f) linking the selected concepts as entities in a knowledge module for performing analysis thereon; and

g) analyzing the knowledge module of linked entities.

2. The method of claim 1, wherein the step of collecting a plurality of concepts includes the step of collecting concepts from a plurality of source classification schemes.

3. The method of claim 1, further comprising the step of uniquely identifying each of the plurality of concepts.

4. The method of claim 1, wherein the step of maintaining a list of the collected plurality of concepts includes the step of creating a concept catalog for storing therein the list of the collected plurality of concepts.

5. The method of claim 1, wherein the step of identifying a location of data pertaining to each of the plurality of concepts includes the further step of maintaining a list of locations of data pertaining to each of the plurality of concepts.

6. The method of claim 5, wherein the list of locations is maintained in a metadata database.

7. The method of claim 6, wherein the metadata database indexes the data pertaining to each of the plurality of concepts by the concept identifier associated with each of the plurality of concepts.

8. The method of claim 1, wherein the step of linking the selected concepts as entities in a knowledge module for performing analysis thereon includes specifying semantic links or data links between the entities.

9. The method of claim 1, further including the step of identifying a method for retrieving data pertaining to each of the plurality of concepts from the location of the data pertaining to each of the plurality of concepts.

10. The method of claim 9, wherein the step of identifying a method for retrieving data pertaining to each of the plurality of concepts comprises the further step of maintaining a list of methods for retrieving data pertaining to each of the plurality of concepts.

11. The method of claim 10, wherein the list of methods is maintained in a metadata database.

12. A computer implemented method of analyzing information, comprising the steps of:

a) collecting in a concept catalog a plurality of concepts from a plurality of disparate classification schemes;

b) uniquely identifying each of the plurality of concepts with a concept identifier and attaching a textual label;

c) identifying a plurality of characteristics for data pertaining to each of the plurality of concepts;

d) maintaining the plurality of characteristics for data pertaining to each of the plurality of concepts in a metadata database, indexed by the concept identifier associated with each of the plurality of concepts;

e) selecting concepts from the concept catalog for inclusion as entities in a knowledge module;

f) linking the entities in the knowledge module for performing analysis thereon; and

g) analyzing the knowledge module.

13. A computer program product comprising a computer usable medium having computer readable program code means embodied therein for causing analysis of information, the computer readable program code means in the computer program product comprising:

a) computer readable program code means for causing a computer to collect a plurality of concepts;

b) computer readable program code means for causing a computer to uniquely identify each of the plurality of concepts with a concept identifier;

c) computer readable program code means for causing a computer to maintain a list of the collected plurality of concepts;

d) computer readable program code means for causing a computer to identify a location of data pertaining to each of the plurality of concepts;

e) computer readable program code means for causing a computer to select concepts from the list of collected plurality of concepts;

f) computer readable program code means for causing a computer to link the selected concepts as entities in a knowledge module for performing analysis thereon; and

g) computer readable program code means for causing a computer to analyze the knowledge module of linked entities.

14. The computer program product of claim 13, wherein the computer readable program code means for causing a computer to collect a plurality of concepts includes the computer readable program code means for causing a computer to collect concepts from a plurality of source classification schemes.

15. The computer program product of claim 13, further comprising computer readable program code means for causing a computer to uniquely identify each of the plurality of concepts.

16. The computer program product of claim 13, wherein the computer readable program code means for causing a computer to maintain a list of the collected plurality of concepts includes the computer readable program code means for causing a computer to create a concept catalog for storing therein the list of the collected plurality of concepts.

17. The computer program product of claim 13, wherein the computer readable program code means for causing a computer to identify a location of data pertaining to each of the plurality of concepts includes computer readable program code means for causing a computer to maintain a list of locations for data pertaining to each of the plurality of concepts.

18. The computer program product of claim 17, wherein the list of locations is maintained in a metadata database.

19. The computer program product of claim 18, wherein the metadata database indexes the data pertaining to each of the plurality of concepts by the concept identifier associated with each of the plurality of concepts.

20. The computer program product of claim 13, wherein the computer readable program code means for causing a computer to link the selected concepts as entities in a knowledge module for performing analysis thereon includes computer readable program code means for causing a computer to specify semantic links or data links between the entities.

21. The computer program product of claim 13, further comprising computer readable program code means for causing a computer to identify a method for retrieving data pertaining to each of the plurality of concepts from the location of the data pertaining to each of the plurality of concepts.

22. The computer program product of claim 21, wherein the computer readable program code means for causing a computer to identify a method for retrieving data pertaining to each of the plurality of concepts comprises computer readable program code means for causing a computer to maintain a list of methods for retrieving data pertaining to each of the plurality of concepts.

23. The computer program product of claim 22, wherein the list of methods is maintained in a metadata database.

24. A program storage device readable by a machine, embodying a program of instructions executable by the machine to perform steps for analyzing information, the steps comprising:

a) collecting in a concept catalog a plurality of concepts from a plurality of disparate classification schemes;

b) uniquely identifying each of the plurality of concepts with a concept identifier and attaching a textual label;

c) identifying a plurality of characteristics for data pertaining to each of the plurality of concepts;

d) maintaining the plurality of characteristics for data pertaining to each of the plurality of concepts in a metadata database, indexed by the concept identifier associated with each of the plurality of concepts;

e) selecting concepts from the concept catalog for inclusion as entities in a knowledge module;

f) linking the entities in the knowledge module for performing analysis thereon; and

g) analyzing the knowledge module.

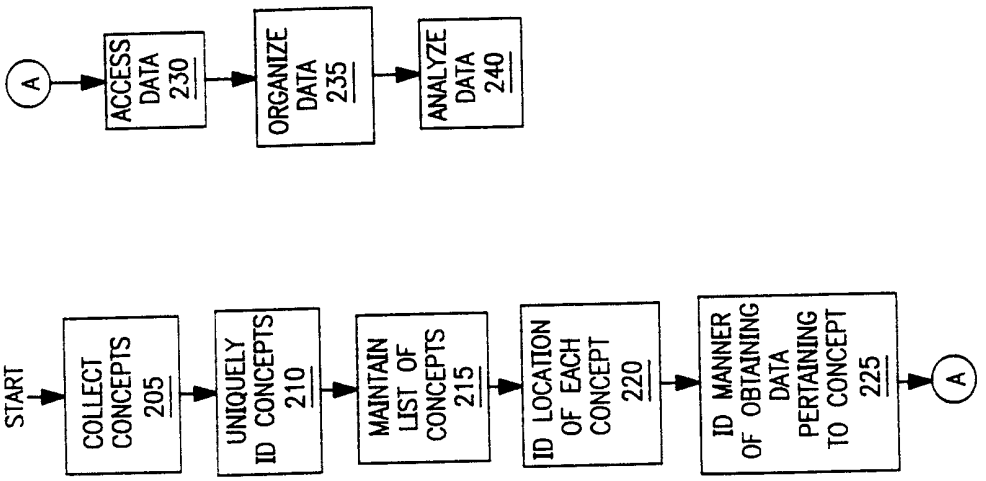

START
COLLECT CONCEPTS 205
UNIQUELY ID CONCEPTS 210
MAINTAIN LIST OF CONCEPTS 215
ID LOCATION OF EACH CONCEPT 220
ID MANNER OF OBTAINING DATA PERTAINING TO CONCEPT 225
A
A
ACCESS DATA 230
ORGANIZE DATA 235
ANALYZE DATA 240

FIG. 2

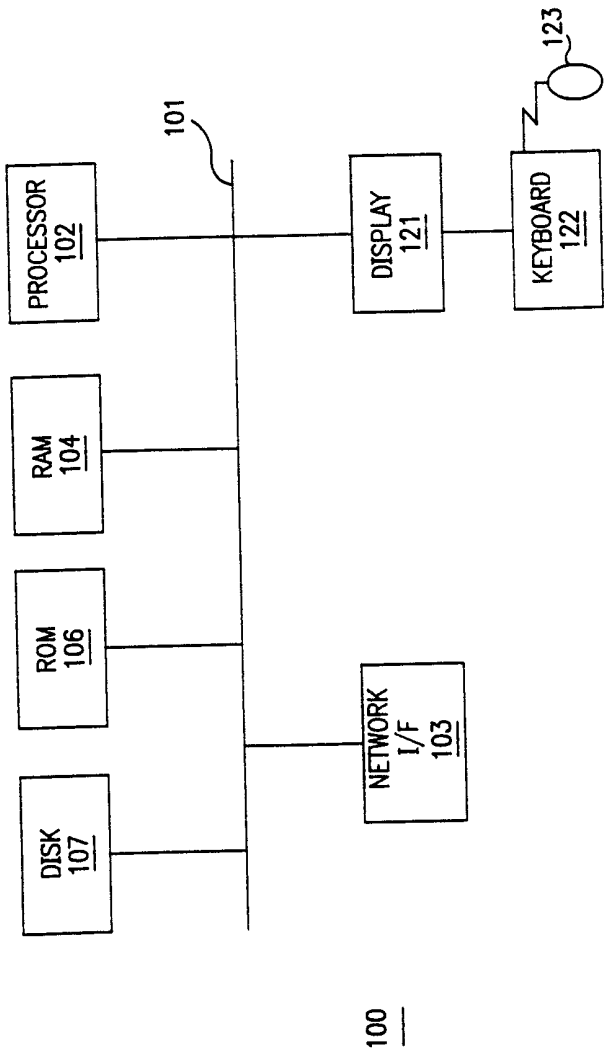

100
101
PROCESSOR 102
RAM 104
ROM 106
DISK 107
NETWORK I/F 103
DISPLAY 121
KEYBOARD 122
123

FIG. 1

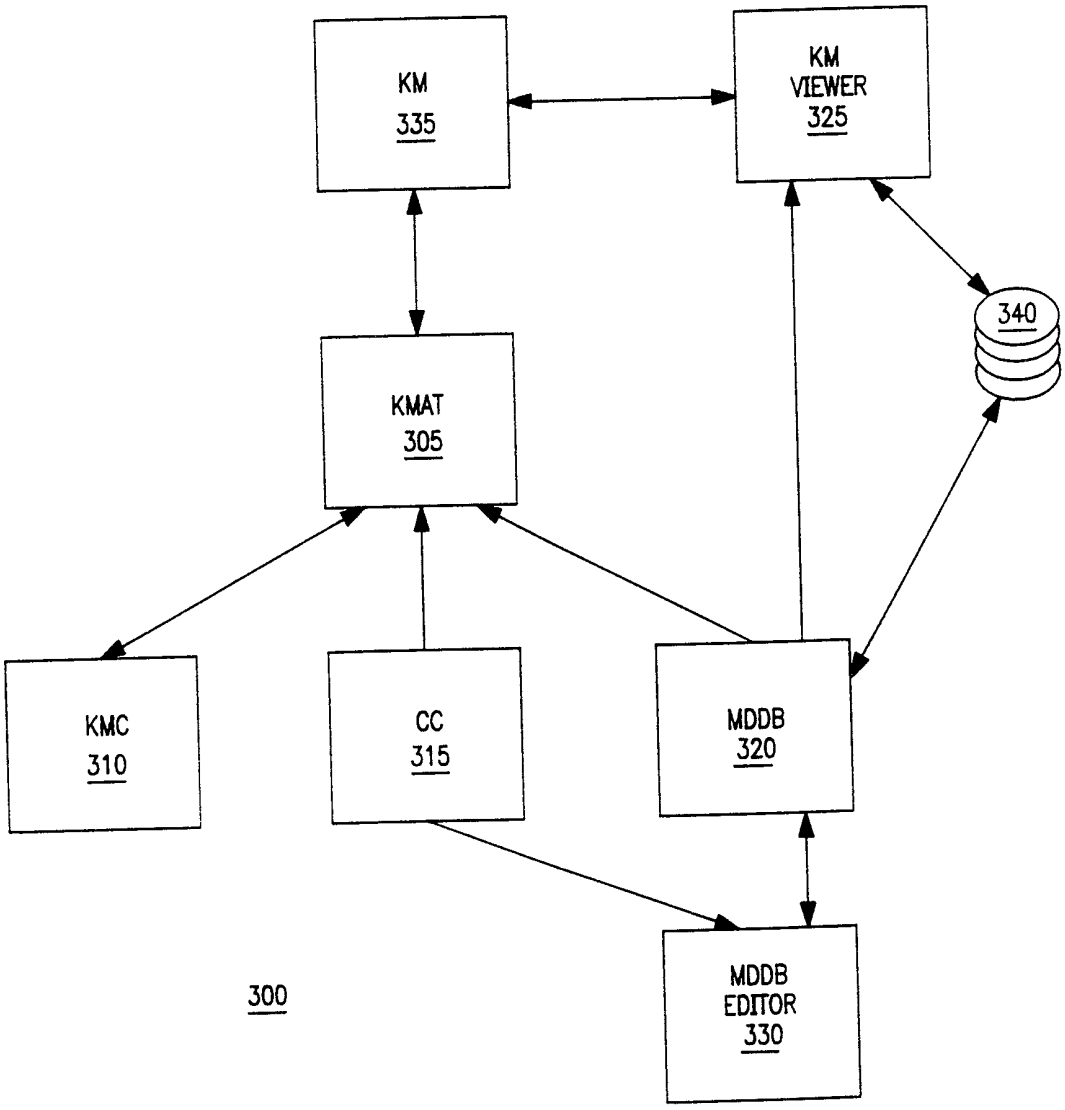

KM
335
KM
VIEWER
325
340
KMAT
305
KMC
310
CC
315
MDDB
320
MDDB
EDITOR
330
300

FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/09741

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G06F 17/30
US CL :707/1,4,5,102

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 707/1,4,5,102

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

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

APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,673,428 A (HIRAKAWA) 30 September 1997, col.3 lines 20-49; note figures 1,2,3,11, and 16	1-26
Y	US 5,799,268 A (BOGURAEV) 25 August 1998, abstract, figure 6, col.59 line 45 through col.60 line 30	1-26
Y	US 5,123,103 A (OHTAKI et al) 16 June 1992, abstract figure 1 and 2	1-26
Y,P	US 5,887,120 A (WICAL) 23 March 1999, abstract, figure 2, 5 and 6	1-26

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents

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

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
09 AUGUST 1999	31 AUG 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer PAUL V. KULIK James R. Matthews Telephone No. (703) 308-3831

Form PCT/ISA/210 (second sheet)(July 1992)★