PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G06F 17/30	A1	(11) International Publication Number: WO 00/08570
		(43) International Publication Date: 17 February 2000 (17.02.00)

(21) International Application Number: PCT/GB99/02495

(22) International Filing Date: 30 July 1999 (30.07.99)

(30) Priority Data:
98306264.7 5 August 1998 (05.08.98) EP

(71) Applicant *(for all designated States except US)*: BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DAVIES, Nicholas, John [GB/GB]; 10 Spindle Wood, High Woods, Colchester, Essex, CO4 4SX (GB). STEWART, Robert, Scott [GB/GB]; Hillbend Cottage, 20 Blind Lane, Coleby, Lincoln, Lincolnshire LN5 0AL (GB).

(74) Agent: DUTTON, Erica, Lindley, Graham; BT Group Legal Services, Intellectual Property Dept., Holborn Centre, 8th Floor, 120 Holborn, London EC1N 2TE (GB).

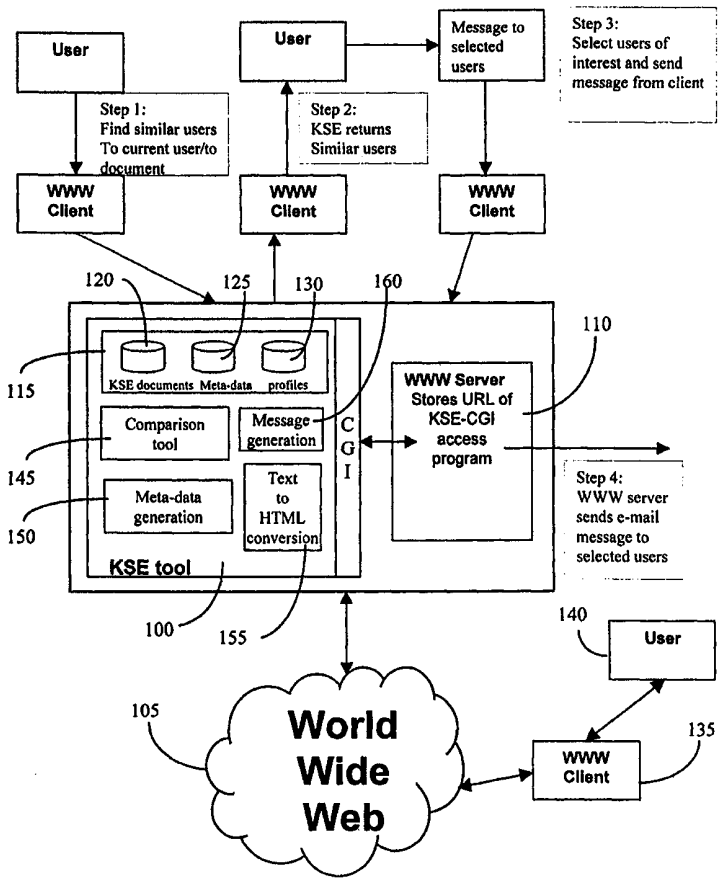
(81) Designated States: AU, CA, SG, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: INFORMATION ACCESS

(57) Abstract

An apparatus (100) is provided to enable information sets stored in an information system and accessible over a communications network to be shared among a group of users. The apparatus (100) is provided with data storage to store a user profile for each user, each user profile comprising at least one keyword expressing an interest of the respective user, and for storing meta–information relating to information sets stored in the information system. The apparatus is provided with means to enable a first user to input an information set, to generate meta–information for that information set, to compare the generated meta–information with stored user profiles of the other users and thereby to alert the first user to the identity of one or more users having a shared interest in the input information set. The apparatus (100) is also enabled to compare the user profile of a first user with profiles of other users to enable the first user to identify other users having similar interests and to enable the first user to send an alert message to one or more of those selected users.



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		

INFORMATION ACCESS

The present invention relates to information access and finds particular application in sharing of information among a group of users.

The Internet is an example of a distributed file storage and retrieval system, being a multimedia computer communications network built on world-wide telephone and data networks. Over 100,000 servers of various types are presently connected to the Internet providing a publicly accessible distributed data store. Data may be stored on a server in a form accessible using an Internet communication protocol called the "HyperText Transfer Protocol" (HTTP). A server storing and making data available in this form is known as an "HTTP server". Data files stored on HTTP servers and accessible by means of HTTP are known as "web pages" which together form the "World-Wide Web", or simply the "Web". Web pages are written using a special Web language called HyperText Mark-up Language (HTML) that includes a facility to create links to other pages on the Web, as appropriate, and enables a user to navigate through information on the Web by means of such links. Information held on the Web is accessible to anyone having a computer connected to the Internet and with an interest in accessing it.

An HTTP Uniform Resource Locator (URL) has been adopted as a Web standard to provide a consistent international naming convention to uniquely identify the location of any Web resource, including for instance documents, programs, sound and video clips. The HTTP enables URL-identified files (web pages) to be located and transferred for reproduction at user equipment connected to the Internet. Underlying transport protocols, primarily TCP/IP, enable connections to be established, between an Internet user and a Web server for example, for the intercommunication of data.

Internet users may access information on the Web using proprietary Web browser products running on personal computers (PCs) or workstations linked to the Internet. Web browsers communicate with Web resources using standard Internet protocols to download selected web pages, interpret embedded HTML commands inserted at the time of mark-up by web page authors and, if appropriate, display those pages graphically. In particular, an HTML command may be embedded within a web page to create a link, by means of a URL, to another web page. Browsers provide a convenient mechanism for displaying the existence

of such links in a displayed web page and for enabling a user to select a particular displayed link and automatically retrieve the referenced page.

There are now estimated to be more than 60 million documents on Web servers world-wide. Every day millions of people trawl the Internet for information using any one of a dozen or more different search tools.

Accompanying this, there has been increasing amounts of information stored in private networks whose operation is based on Internet Protocols (IP). These private networks are commonly referred to as intranet networks.

One of the main advantages of an intranet for a corporation is the ease with which information, that was previously localised to individual departments, can be accessed from across the organisation as a whole.

Information management tools such as the JASPER agent detailed in British Telecommunications plc's co-pending application WO96/23265 provide a knowledge sharing facility whereby users may be alerted to the storage or identification of documents or other information sets according to whether those users' interests, defined by user profiles, match the content of the stored document. However, while users may be alerted to the existence of documents of potential interest, there is no facility for interaction between users of the information system.

According to a first aspect of the present invention there is provided an apparatus, for accessing sets of information stored in an information system and accessible by means of a communications network, the apparatus having:

i) an input for receiving a set of information selected by a first user;

ii) data storage, or means to access data storage, for storing one or more user profiles, each said user profile comprising at least one predetermined keyword;

iii) generation means, triggerable to generate at least one set of meta-information from the set of information received at the input, said meta-information including at least a pointer for the set of information, and to store said at least one set of meta-information in said data storage;

iv) comparison means for comparing at least one of said one or more user profiles with said at least one set of meta-information and for identifying, in dependence upon the results of said comparison, a user having a profile similar to said at least one set of meta-information; and

v) alerting means to alert said first user to the identity of said user identified by said comparison means.

According to this first aspect, a user submitting an information set may be alerted to the identity of other users interested in that information set and potentially sharing an interest with the user submitting that information set.

Preferably, the comparison means are operable to compare a user profile associated with said first user with at least one further user profile and thereby to identify a user having a similar user profile to that of said first user. In this way, a user may carry out a search for other users having certain interests in common on the basis of their users profiles.

In a further embodiment, the apparatus may include selecting means to enable said first user to select one or more of said identified users and to generate an alert message for sending to said one or more selected users. Users identified to the first user may include those having an interest in an information set submitted by the first user or those identified as a result of comparing user profiles.

An information system operating over the Internet for example, may be arranged to operate only with HTML documents. According to a further embodiment of the present invention, where stored sets of information conform to a first predetermined format, the apparatus includes conversion means to enable a set of information received at the input in a format other than said first predetermined format to be converted into said first predetermined format and stored in said data storage. In this way, information such as e-mails and other electronic documents, can be shared.

When storing an e-mail or an electronic document, on a Web site or a Web page, the document or e-mail needs to have some measure of relevance to the pre-existing content, and possibly the structure, of the Web page or Web site on which it is stored. If it only has marginal relevance, or no relevance at all, to the content or structure of the Web page or Web site on which it is stored, then it may be difficult for parties interested in the information to locate it.

Hence, when a person decides to share some information by means of an intranet, it usually requires that they either design a new Web page or that they re-design an existing Web site. This ensures that the information is stored in a logical manner on the intranet.

However, such designing, or re-designing of a Web page represents a significant amount of work. This usually results in the information in question remaining stored on a local server or an e-mail system.

However, most e-mail systems and local servers only permit access to the information they store by a closed group of users. Also, many e-mail systems and local servers do not provide access to the information that they store by means of Internet protocols.

Embodiments of the present invention provide an interface between systems that store information in a distributed manner, such as the Internet and corporate intranets, and information that does not fit logically within the pre-existing pages of such systems.

An example of the type of information that may not fit logically within a pre-existing Web page or Web site is an e-mail or a memo. It is known to store e-mails on the Web via news groups. However, news groups have a logical structure whereby e-mails are grouped into related topics.

However, where an e-mail relates to, say, a particular project and where that project has its own Web site, it may not be sufficient to simply store the e-mail somewhere on the Web site. It may be necessary to create a separate Web page for the e-mail so that it can be viewed within a particular context.

Embodiments of the present invention provide functionality for interfacing e-mails and other electronic documents with the Web. This ability represents an improvement to the tool detailed in WO96/23265, referred to above.

The use of profiling can provide a pro-active information sharing scheme whereby users who have previously had an interest in a particular topic are alerted when a new item of information matching that topic is stored within the information access system.

Thus e-mails and other electronic documents can be pro-actively distributed to users whose interests are referenced by a tool or system that embodies the present invention.

In particular the person storing the e-mail need not know every person to whom the information needs to be distributed. Thus an e-mail relating to a particular project can be stored and then shared with future members of the team.

According to a second aspect of the present invention there is provided a method of monitoring stored information sets accessible by means of a

communications network, for the purpose of alerting a first user to the existence of a second user having a shared interest in an information set selected by said first user, the method comprising the steps of:

a) storing a user profile for each user, which profile comprises at least one keyword and an identifier for the user;

b) receiving a set of information selected by said first user;

c) generating a set of meta-information dependent on said received information set;

d) comparing the generated set of meta-information with a stored user profile other than that for said first user and, in dependence upon the result from the comparison, identifying a second user having a user profile similar to said meta-information;

e) transmitting an alert message addressed to the first user comprising at least the identity of said second user.

Embodiments of the present invention will now be described, by way of example only and with reference to the accompanying drawings, of which:

Figure 1 is a schematic representation of an information access system incorporating embodiments of the present invention;

Figure 2 is a flow chart of operation performed by the information access system of Figure 1;

Figure 3 is a flow chart of step 205 of Figure 2 in greater detail;

Figure 4 is a flow chart of step 335 of Figure 3 in greater detail;

Figure 5 is a flow chart showing the steps by which a user may access meta-information;

Figure 6 is a flow chart showing the steps in operation of the system when a user selects a hype-text link;

Figure 7 is a flow chart showing the steps involved in searching for other users having similar user profiles;

Figure 8 is a flow chart showing the steps in identifying other users having an interest in a particular information set.

Figure 1 depicts an information access system, referred to as a Knowledge Sharing Environment tool (KSE tool) 100, that is connected to an information

system that stores information in a distributed manner, such as a corporate intranet or the Internet 105.

The KSE tool 100 has a KSE data store 115 that stores sets of KSE documents 120 (detailed further below), sets of meta-data 125 and sets of user profiles 130.

The KSE documents are generated by the text to HTML conversion application 155 and the meta data 125 is generated by the meta data generation application 150. A message generation application 160 is used for automatically generating e-mail messages that are sent to users 140 of the KSE tool. A comparison application program 145 is used for comparing items of meta data 125 against user profiles 130 and search queries. It may also be used for comparing user profiles 130 against other user profiles 130.

The KSE tool 100 may typically be hosted by a communications terminal that is connected to a communications network and that is accessible by use of software clients, such as Internet and Web browsers 135.

The KSE tool 100 also has access to a Web server 110 by means of a CGI program. The KSE tool 100 uses the Web server 110 to communicate KSE documents 120 and other information to the Web browsers 135. The Web server 110 preferably has the capability to transmit electronic mail to its users 140 using the Simple Mail Transfer Protocol (SMTP).

Accordingly, the KSE tool 100 can be said to exist in a distributed system or a distributed environment, often called a heterogeneous distributed system or environment. Such systems are typically based on a client server architecture. They may have numerous clients with the capability of accessing numerous servers by means of one or more communications networks. The Internet is an example of such a distributed environment.

The present invention however, is not limited to the Internet. It may be implemented in other distributed systems based on other protocols and languages. For examples, Object (as in Object Oriented) based systems that use CORBA (Common Object Request Broker Architecture).

Referring now to Figure 2, an overview of the operation of the KSE tool 100 is provided. In operation, a user 140 may have at step 200 generated or located some information that they determine should be shared between the users

of the KSE tool 100. At step 205, in order to share this information via the KSE tool 100, the user 140 stores it as a KSE document 120.

At step 210 the KSE tool generates meta data in respect of the newly stored KSE document 120. And at step 215, the KSE tool 100 compares this meta-data 125 against each of the user profiles 130 that the KSE tool 100 has stored in the KSE data store 115.

Meta-data is a term used to refer to a data set that describes information, i.e. information about information. A simple set of meta-data would be a summary of a web page combined with the URL of the web page. Some other types of information that may also form meta-data include, details on when a Web page was last updated, who it was that updated it etc and any previous URLs that the page may have been stored under.

At step 220, those users whose profile 130 matches the meta data 125 of the KSE document 120 are notified that it has been stored within the KSE tool 100. This notification may take place, for example, by e-mail or on the opening screen of the KSE tool at their next KSE session.

In this way, use of user profiles 130 allows the information to be easily shared amongst users of the KSE tool without the need either to generate a new Web page or to re-design an existing Web page. An additional advantage is that the user profiles 130 enable the KSE document 120 to be shared amongst users 140 of the KSE tool 100 without the user 140 who is storing the KSE document 120 needing to know who might have an interest in it. This also enables the information to be shared with future users of the KSE tool 100 who may not have an account when the information is stored. Similarly, it allows the information to be shared with users whose profile may change at some future point in time so as to match the meta-data 125 now being stored.

Typically the sort of information that will be stored as a KSE document 120 is e-mail messages and electronic documents such as memos, reports, letters of advice and other information that may not fit logically within the pre-existing contents or structure of a Web page or a Web site.

Referring now to Figure 3, further detail of generating KSE documents as detailed at step 205 of Figure 2 is provided. As detailed above, at step 205, a user 140 submits an item of information to the KSE tool 100. At step 305 a new text file is created within the KSE data store 115. This text file may be generated

through the use of a text box on the Web browser 135. The browser 135 may then submit the text file to the KSE tool 100.

At step 310, the KSE tool 100 stores the text file in the KSE store 115 by assigning it a number, which is one greater than the number of the previous text file stored. Alternatively the text file may be stored within an area of the KSE store 115 that is dedicated to the user 140 storing the text file. In this case, the number assigned to the text file is one greater than the number assigned to previous KSE document 120 stored in that user's area.

At step 315, the KSE tool 100 places a Hyper Text Mark-up Language (HTML) framework around the newly stored text file. Typically, the HTML framework is HTML tags indicating the title of the page and tags indicating the body of the text. This framework may also include under an HTML meta tag, a comment detailing the time and date that the text file was stored in the KSE tool 100 and an identifier for the user who submitted the information. The HTML framework preferably includes information on the appearance of the KSE document 120 on a Web browser 135. This HTML framework enables the text file to be passed from the KSE tool 100 to the Web server 110 and then on to a Web browser 135, using Internet protocols.

At step 320, once the KSE document 120 has been generated and stored, the KSE tool 100 then generates meta data 125 about the KSE document 120 and then at step 325 stores this meta data 125 in the KSE data store 115.

This meta data 125 typically includes a URL pointing to a location on the Web server 110 where the KSE document 120 can be accessed.

The meta data may also contain a summary of the information in the KSE document 120. Such a summary may be generated automatically by a text processing tool such as the ConText™ tool from Oracle. Alternatively the summarisation tool called Prosum detailed in "Davies, NJ, Weeks, R (1998) ProSum Profile based text summarisation, First Automatic Text Summarisation Conference (SUMMAC-1), May 1998, Virginia USA" may be used.

The meta data may also contain a list of key words and key phrases appearing in the document. Such a list of key words and key phrases may be extracted by text processing tools such as those listed above in relation to tools for generating summaries of text documents.

The meta data may also include details such as the person who requested that the information be stored as a KSE document 120 and the date and time that it was stored in the KSE tool 100. The meta data may also contain an annotation about the original document. This annotation may be entered by a user when the document is stored or by subsequent readers of the document. Such annotations usually provide a clue as to why the document was selected for storage. For example, where the document is relevant to a particular project, the annotation may simply be the name of the project concerned.

The meta data, having an embedded link to the relevant KSE document 120 is the means by which a user 140 of a Web browser gains information about the KSE document 120 concerned and also the URL of the KSE document 120 concerned.

In this sense the meta data generated by the KSE tool 100 can be seen as an index of the KSE documents 120 that are referenced by the KSE tool 100.

Once the text file has been stored and an HTML framework placed around it, the text file may be referred to as a KSE document 120. The file name of the KSE document 120 may be used as a parameter that can be passed to a KSE-CGI access program stored in the CGI bin of the Web server 110.

The KSE-CGI access program may be a standard form of CGI program used for accessing data bases, such as the KSE data store 115, via the Internet. With such an arrangement, a standard URL will designate the KSE-CGI access program. The file name of the KSE document will be passed as a parameter URL.

The KSE-CGI access program may operate in accordance with standard CGI techniques for retrieving data from a data base into a Web server.

The KSE document 120 may be retrieved into the Web server 110, by passing a KSE document 120 file name, to the KSE-CGI access program, as a parameter within a URL. The URL containing the file name parameters will be the URL of the KSE-CGI program. The parameters may be passed within the URL in a manner similar to passing query strings to search engines.

Once a KSE document 120 has been loaded by the KSE-CGI program into the Web server, it may then be loaded to the browser that requested it.

The KSE-CGI access program provides a common interface facility between the KSE tool 100 and the Web 105. When the KSE tool 100 is initially accessed, it may present information to a browser 135 as a list of meta data 125.

Preferably the meta data 125 is presented in chronological order from the most recently generated item of meta data 125 down to the first item of meta data that was created. This enables the KSE tool to present the most recently stored information to the user when they next log on to the KSE tool 100.

The KSE-CGI access program and the meta-data 125 operate together to form a Web site where the content of the information presented and the structure with which it is presented can change dynamically.

For example, the meta data allows KSE documents with similar information content to be referenced against each other. This, for example, allows e-mails and other electronic documents to be grouped together according to their project name. Hence a user need only do a search on the project name to retrieve the relevant information stored in the KSE tool 100.

This referencing of new KSE documents 120 against other KSE documents 120 allows new items of information to be added to the KSE tool 100 without the need for creating a new Web page or for amending an existing Web page.

This referencing also assists with reducing problems of information overload as the new KSE document 120 can viewed in the context of other relevant KSE documents 120.

This structure of the KSE tool 100, with a KSE-CGI access program separating KSE documents 120 from HTML files on the Web server files, has a further advantage of increasing the ease with which the KSE tool 100 can be ported from one server to another.

The KSE tool also provides a mechanism whereby users 140 of the KSE tool 100 may be automatically notified when the KSE tool 100 generates an item of meta data that matches their profile. This is achieved by comparing newly generated items of meta data 125 against the user profiles 130 held in the KSE store 115.

Referring now to Figure 4, the following details how the user profiles 130 are compared against newly created KSE documents 120.

At step 400, an item of meta data 125 is generated by the KSE tool 100. At step 405 the KSE tool 100 accesses the first user profile 130 stored in the user's KSE data store 115. At step 410, the KSE tool 100 compares the item of meta-data 125 against the user profile using a vector space algorithm. More detail on the vector space algorithm is provided below. If at step 415, a positive

comparison is detected, then at step 420 an alert message, which may be an e-mail message, is generated by the KSE tool 100 and at step 425 the e-mail address of the user concerned is retrieved by the KSE tool 100 from the users profile information 130 that is stored in the KSE store 115.

At step 430 the alert message along with the user's 140 e-mail address is sent to the Web server 110 which then combines the message and the e-mail address into an e-mail. At step 440, the Web server 110 transmits the alert message to each user concerned.

After this, at step 445, the KSE tool 100 tests whether all of the user profiles stored in the KSE data store 115 have been compared against the newly stored item of meta data 125. If they have, then the process ends. If they have not, then the KSE tool 100 loops around to commence from step 405 above, where it accesses the next user profile 130 stored in the KSE store 115.

Alternatively, when a user amends their user profile, by adding one or more key words or key phrases, the KSE tool 100 may compare the new profile against all of the meta-data 125 stored in the KSE store 115. The user is then alerted to any items of meta-data 125 that match the changed profile 130. Similarly, when a new user 140 is provided with access to the KSE tool 100 their profile 130 is compared against the meta-data 125 in a manner similar to that of Figure 4

A user profile 130 is typically a list of key words and key phrases that define topics that a user 140 is interested in. Further detail on user profiles may be found in British Telecommunications plc's co-pending application WO96/23265 referred to above.

The above description refers to the KSE tool 100 referencing documents generated by its users. However, the KSE tool 100 need not be limited to this. The KSE tool 100 may also reference Web pages and other Internet and intranet sites.

This may be achieved by passing to the KSE tool 100 a URL (Internet address) of a Web page that a user 140 deems worthy of being shared with other users 140 of the KSE tool 100.

In response to receiving the URL, the KSE tool 100 causes the Web server 110 to access the Web page in question and to generate meta data in respect of it. This meta data is generated in the same or a similar manner as discussed above

in relation to the KSE tool and typically comprises the URL in question along with a summary of the Web page held by the URL and a set of key words for the Web page in question.

As stated above, the meta data for each document referenced by the KSE tool 100 contains a list of the key words and key phrases that have been extracted from the document in question. Each key word or key phase may have a number associated with it that corresponds to the number of times that the key word or key phrase in question occurs in the document from which it was extracted.

These numbers are then copied into their corresponding entry in a T x D term document matrix **M**. The matrix **M** represents a set T of all the terms t_i appearing in all the sets of meta data held by the KSE tool 100. The matrix **M** also represents the set D of all the documents d_i referenced by the KSE tool 100. The sets T and D are set against each other to form the matrix **M**. The value of each entry in the matrix **M** is obtained from the list of key terms and key phrases generated for the meta data, i.e. the term document matrix identifies the number of times each term t_i appears in a document d_i.

The term document matrix **M** may be searched against both user profiles and key word search queries, so as to locate documents referenced by the KSE tool 100 that match either the user profile or the search query. This is because user profiles and search queries are similar in that they both consist of key words and key phrases.

To compare a user profile against the matrix **M**, a vector space matching and scoring algorithm may be employed. An example of such an algorithm is detailed in Salton, G "Automatic Text Processing", Addison – Wesley, Reading, Mass., USA, 1989. It measures the similarity between a user profile p and a document d according to:

$$sim(p,d) = \sum_{i=1,n} \left(t_{ip} * t_{id}\right) / \sqrt{\left[\sum_{i=1,n} t_{ip}^2 * \sum_{i=1,n} t_{id}^2\right]}$$

Where t_{ip} is the i^{th} term or phrase of user profile p;

where t_{id} is the i^{th} key term or key phrase of document d; and

where n is the number of unique terms in the combined profiles of document d and the profile p of the user in question.

The matrix **M** is of indefinite size as it cannot be determined how many documents or terms it will reference. For this reason, the matrix **M** may be created using a dynamic array, linked lists or similar types of data structures.

This vector space matching and scoring algorithm may also be employed in the identification of users of the KSE tool 100 who have similar profiles. In this case, a term user profile matrix **U**, rather than the term document matrix **M**, is created. The term user profile matrix **U** is a matrix of the set of all the terms **T** in all of the user profiles held by the KSE tool 100. This set of terms **T** is set against the set of all of the user profiles **P** of the KSE tool 100. The entries in the matrix **U** may be either a 1 or a 0. Alternatively, the value of the entry may indicate the user's level of interest in the term.

Using the vector space scoring and matching algorithm against the term user matrix ***U*** generates a measure of the similarity between the profiles of two users U_1 and U_2. In this instance, the algorithm is expressed as:

$$sim(u_1, u_2) = \sum_{i=1,n} (t_{i1} * t_{i2}) / \sqrt{\left[\sum_{i=1,n} t_{i1}^2 * \sum_{i=1,n} t_{i2}^2\right]}$$

Where t_{i1} is the i^{th} term or phrase of user U_1 profile;

where t_{i2} is the i^{th} term or phrase of user U_2 profile; and

where n is the number of unique terms in the combined profiles of users U_1 and U_2.

Referring to Figure 7, a flow diagram is provided showing the steps in using the term user matrix ***U*** to identify other users of the KSE 100 having a similar profile of interests to a first user U_1 for example. Having generated the term user matrix ***U*** at step 700, at step 710 the above algorithm is run against the matrix ***U*** to determine, in turn, a measure of the similarity between the profile of user U_1 and that of each of the other users. At step 720, those users having the most similar profiles to that of user U_1 are identified and user U_1 is notified of the identity if

each such user. At step 730, user U_1 is able to select one or more users from those notified such that the KSE 100 is triggered to generate a message for sending to each selected user, for example as an introduction from user U_1 and to inform the selected user of their shared interests.

Referring to Figure 8, a flow diagram is presented showing the steps in operation of the KSE 100 in enabling a user U_1 to identify any other users whose user profile matches meta data generated from a set of information submitted or selected by the user U_1. Following receipt by the KSE 100 of the information set at step 800, at step 810 the KSE 100 generates a set of meta-information for the received information set. At step 820 a matching algorithm, such as the vector space scoring and matching algorithm described above, is applied to compare the generated meta-information with the user profiles of each of the users of the KSE 100. At step 830, on the basis of the matching algorithm results, those users having user profiles most similar to the generated meta-information, and hence with a likely interest in the information set submitted by user U_1, are identified and their identities notified to user U_1. At step 840, user U_1 is able to select one or more users from those notified such that the KSE 100 is triggered to generate a message for sending to each selected user, for example as an introduction from user U_1 and to inform the selected user of their shared interests.

The KSE tool 100 however is not limited to simply referencing KSE documents 120. The KSE tool 100 may also reference other Web pages. To do this, a user 140 submits a URL to the KSE tool. On receipt of such a URL, the KSE tool 100 instructs the Web server to access the Web page concerned. Meta-data for the page is then constructed in a manner similar to that for KSE documents 120. i.e. a title is stored, a summary is generated, key words are extracted, date and time of storage in the KSE tool is recorded as is the name of the user storing the information, etc.

The meta data for Web pages is preferably indistinguishable from the meta data stored for KSE documents 120. This is to enable e-mails and other electronic documents stored as KSE documents 120 to be shared with other users of the KSE tool indistinguishably from Web pages.

By enabling KSE documents 120 to be shared indistinguishably from other Web pages, the KSE tool 100 provides a seamless Web based interface between

traditional Web documents and other electronic documents and e-mails that traditionally have not been shared by use of the Web.

Referring now to Figure 5 the following details how the KSE tool 100 retrieves and displays meta data when a user 140 first accesses the KSE tool 100.

At step 500 the KSE-CGI access program receives a HTTP Get command from a user's 140 browser 135 requesting that the title page of the KSE tool 100 be down loaded to their Web browser 135.

At step 505 the KSE-CGI access program accesses and retrieves from the KSE data store 115 the item of meta data that was last generated. At step 510 the KSE-CGI access program extracts the URL that is stored as part of the meta data and at step 515 it also extracts the title of the information from the meta data 125.

At step 520 KSE-CGI access program combines the title information and the URL so as to form a HTML Hyper Text Link.

At step 525, the HTML Hyper Text Link is uploaded by the Web server 110 to the browser 135 that transmitted the HTTP Get request at step 500 above. Select items of meta data may also be uploaded to the Web browser 135 at the same time as the hyper Text Link. These may be, for example, the date that the meta data was generated. The user 140 causing the meta data 125 to be generated and any comment that that user 140 stored with the meta data.

At step 530, the KSE-CGI access program tests to see whether or not the first item of meta data 125 that was generated has been uploaded to the browser 135. If it has not, then at step 535, the KSE-CGI access program accesses the item of meta data 125 that was generated immediately before the item of meta data uploaded to the Web browser 135 at step 525 above. The KSE-CGI access program then loops around to repeat steps 510, 515, 520 and 525 detailed above.

Alternatively, step 530 may test to see whether or not a set number of items of meta data have been uploaded to the browser 135. Where there are a large number of items of meta data, this option may be preferred as it enables a small number of items of meta data to be displayed with out crowding the screen of the Web browser 135. An additional control loop may enable the succeeding items of meta data to be uploaded, at a users request, at a later stage.

When step 530 returns a positive result, that all of the items of meta data have been uploaded to the Web browser, then at step 540, the CGI-KSE program

uploads to the browser 135 as KSE navigation tool bar. This KSE navigation tool bar is preferably up loaded as one of two HTML frames. The other HTML frame displays the meta data.

The KSE tool bar preferably displays HTML buttons that enable a user to navigate within the KSE tool 100. The use of two frames, one with a KSE tool bar enables a user to remain in the KSE environment when viewing a Web page stored outside of the KSE tool 100.

The process detailed in relation to Figure 5 is not limited to when a user first access the KSE tool 100. A similar process for displaying search results, subsequent to a user searching the meta data 125, may also be used.

Referring now to Figure 6, the following details the operation of a browser and the KSE-CGI access program when a user selects a hyper text link that has been uploaded to their browser 135 at step 525 of Figure 5.

At step 600, when a user 140 selects a hyper text link, the KSE-CGI access program receives from the users browser 135 a HTTP Get request for a HTML frame set. The HTML frame set contains two frames. The first displays a KSE tool bar and the second displays a user 140 selected Web page or KSE document 120.

At step 605, the frame set instructs the browser 135 to upload the requested Web page or KSE document into the second HTML frame of the frame set, which is achieved by the browser issuing further HTTP get request for the required web page or KSE document.

If the selected Web page is a KSE document 120 then a second HTTP Get request is sent to KSE-CGI access program. The KSE-CGI access program will access the KSE document 120 into the Web server 110 which then uploads it to the second frame on the browser 135.

When the requested Web page is returned to the Web browser 135 it is uploaded into the second HTML frame on the browser display.

At step 610, the frame set requests the KSE-CGI access program to upload a KSE tool bar, referred to as a KSE feedback tool bar to the first HTML frame.

This KSE feedback tool bar enables the selected Web page or KSE document 120 to be viewed from within the KSE tool 100. It also allows KSE functionality to be accessed whilst viewing a selected Web page or KSE document 120.

The KSE feedback tool bar preferably contains HTML buttons that enable a user to indicate their interest in the selected Web page or KSE document 120. This feedback on interest provided may be used to update the users profile. The feedback tool bar also provides a means for gaining access back into the KSE tool 100.

CLAIMS

1. An apparatus, for accessing sets of information stored in an information system and accessible by means of a communications network, the apparatus having:

i) an input for receiving a set of information selected by a first user;

ii) data storage, or means to access data storage, for storing one or more user profiles, each said user profile comprising at least one predetermined keyword;

iii) generation means, triggerable to generate at least one set of meta-information from the set of information received at the input, said meta-information including at least a pointer for the set of information, and to store said at least one set of meta-information in said data storage;

iv) comparison means for comparing at least one of said one or more user profiles with said at least one set of meta-information and for identifying, in dependence upon the results of said comparison, a user having a profile similar to said at least one set of meta-information; and

v) alerting means to alert said first user to the identity of said user identified by said comparison means.

2. An apparatus according to Claim 1, for use by a plurality of users, each of said plurality of users having at least one associated user profile stored in said data storage, wherein the apparatus is triggerable, on activation of said generation means to generate a set of meta-information by said first user, to compare said at least one set of meta-information with at least one user profile associated with a second user and, in dependence upon the result of said comparison, to address an alert message to said first user comprising at least the identity of said second user.

3. An apparatus according to Claim 1 or Claim 2, wherein said comparison means are operable to compare a user profile associated with said first user with at least one further user profile and thereby to identify a user having a similar user profile to that of said first user.

4. An apparatus according to any one of claims 1 to 3, including selecting means to enable said first user to select one or more of said identified users and to generate an alert message for sending to said one or more selected users.

5. An apparatus according to any one of claims 1 to 4, wherein, in use, said stored sets of information conform to a first predetermined format and wherein said apparatus includes conversion means to enable a set of information received at the input in a format other than said first predetermined format to be converted into said first predetermined format and stored in said data storage.

6. An apparatus according to Claim 3, including monitoring means operable to detect a change to the user profile of said first user and to trigger said comparison means to compare the changed user profile with other user profiles stored in said data storage and thereby to identify a user having a similar user profile to the changed profile of said first user.

7. An apparatus according to Claim 1, including monitoring means operable to detect a change to the user profile of said first user and to trigger said comparison means to compare the changed user profile with meta-information stored in said data storage and thereby to alert said first user to a stored information set matching the changed profile.

8. An information access system comprising a plurality of software agents, each agent comprising elements i) to v) inclusive of the apparatus according to claim 1 and each agent being allocated to a different respective user of the system.

9. A method of monitoring stored information sets accessible by means of a communications network, for the purpose of alerting a first user to the existence of a second user having a shared interest in an information set selected by said first user, the method comprising the steps of:

a) storing a user profile for each user, which profile comprises at least one keyword and an identifier for the user;

b) receiving a set of information selected by said first user;

c) generating a set of meta-information dependent on said received information set;

d) comparing the generated set of meta-information with a stored user profile other than that for said first user and, in dependence upon the result from the comparison, identifying a second user having a user profile similar to said meta-information;

e) transmitting an alert message addressed to the first user comprising at least the identity of said second user.

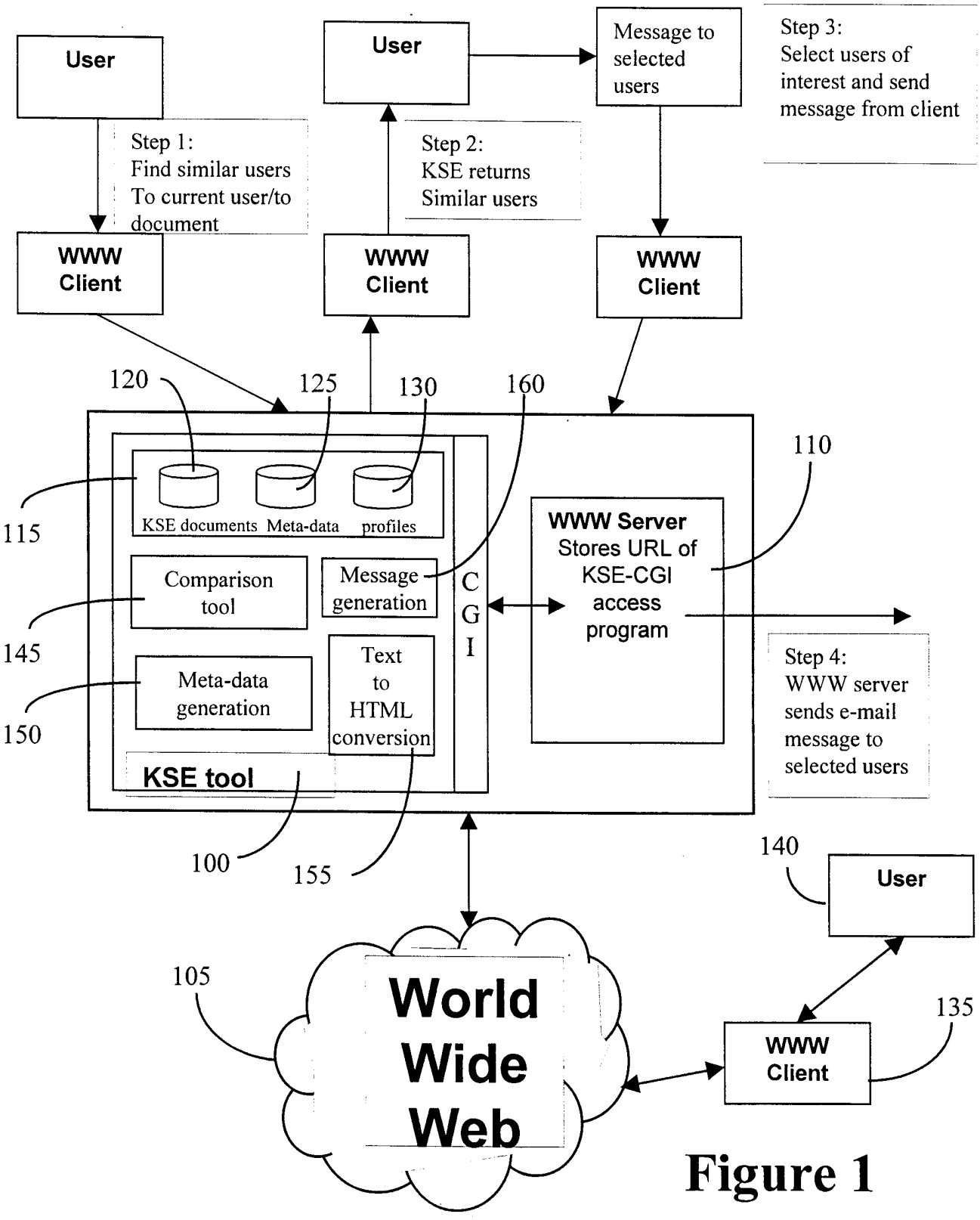

User
User
Message to selected users
Step 3: Select users of interest and send message from client
Step 1: Find similar users To current user/to document
Step 2: KSE returns Similar users
WWW Client
WWW Client
WWW Client
120
125
130
160
110
115
KSE documents
Meta-data
profiles
Comparison tool
Message generation
C G I
WWW Server Stores URL of KSE-CGI access program
145
Text to HTML conversion
Meta-data generation
150
Step 4: WWW server sends e-mail message to selected users
KSE tool
100
155
140
User
105
World Wide Web
135
WWW Client

Figure 1

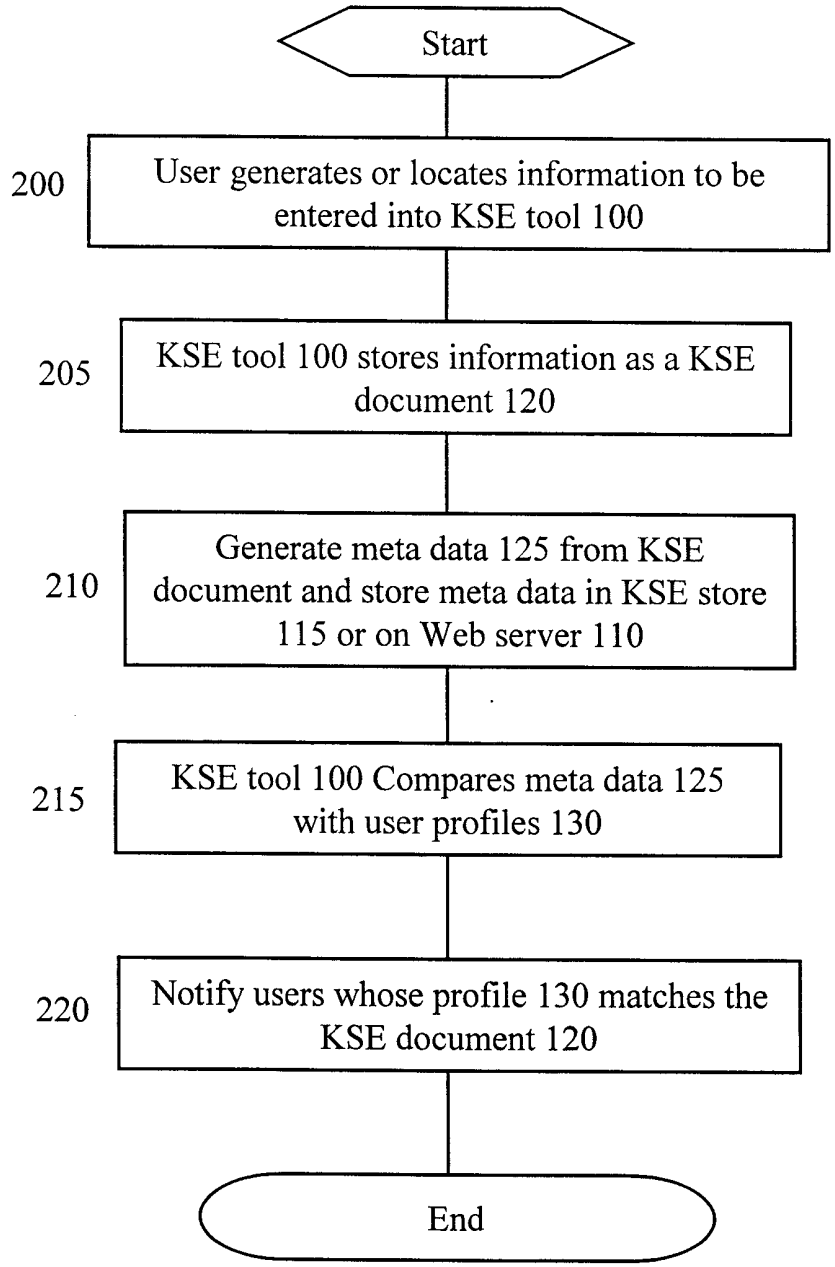

Start
200
User generates or locates information to be entered into KSE tool 100
205
KSE tool 100 stores information as a KSE document 120
210
Generate meta data 125 from KSE document and store meta data in KSE store 115 or on Web server 110
215
KSE tool 100 Compares meta data 125 with user profiles 130
220
Notify users whose profile 130 matches the KSE document 120
End

Figure 2

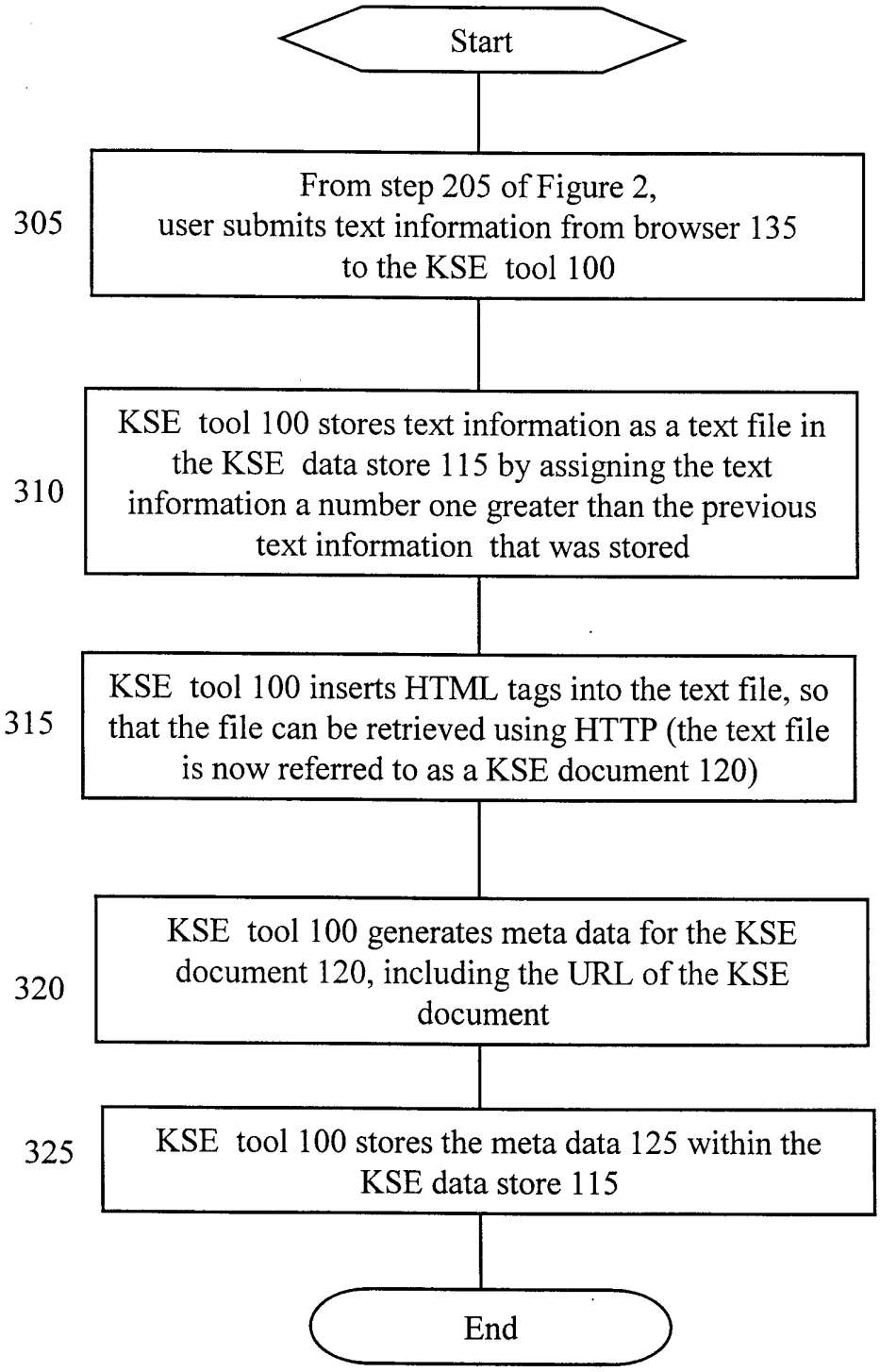

Start
305
From step 205 of Figure 2, user submits text information from browser 135 to the KSE tool 100
310
KSE tool 100 stores text information as a text file in the KSE data store 115 by assigning the text information a number one greater than the previous text information that was stored
315
KSE tool 100 inserts HTML tags into the text file, so that the file can be retrieved using HTTP (the text file is now referred to as a KSE document 120)
320
KSE tool 100 generates meta data for the KSE document 120, including the URL of the KSE document
325
KSE tool 100 stores the meta data 125 within the KSE data store 115
End

Figure 3

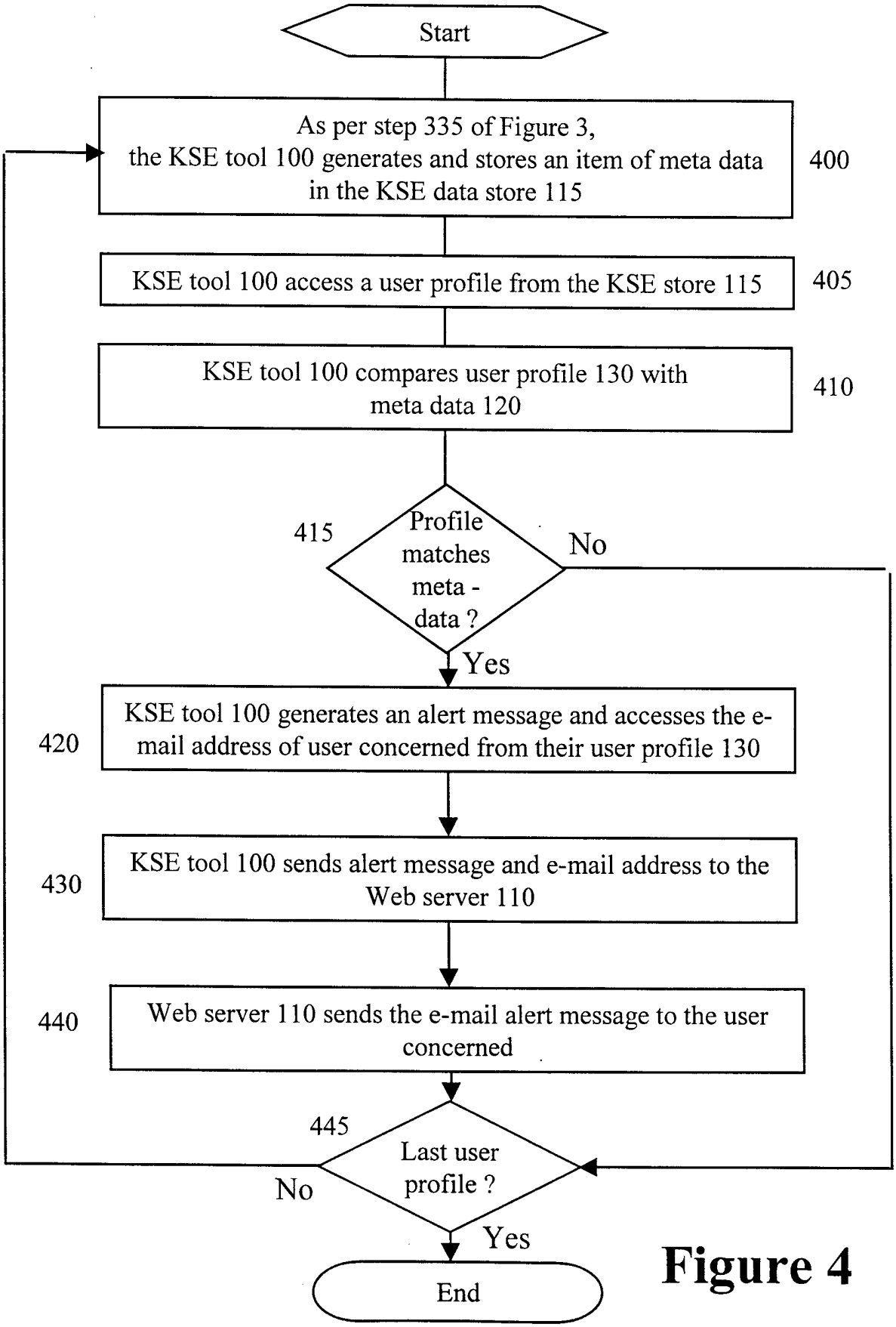

Start
As per step 335 of Figure 3, the KSE tool 100 generates and stores an item of meta data in the KSE data store 115
400
KSE tool 100 access a user profile from the KSE store 115
405
KSE tool 100 compares user profile 130 with meta data 120
410
415
Profile matches meta - data ?
No
Yes
KSE tool 100 generates an alert message and accesses the e-mail address of user concerned from their user profile 130
420
KSE tool 100 sends alert message and e-mail address to the Web server 110
430
Web server 110 sends the e-mail alert message to the user concerned
440
445
Last user profile ?
No
Yes
End

Figure 4

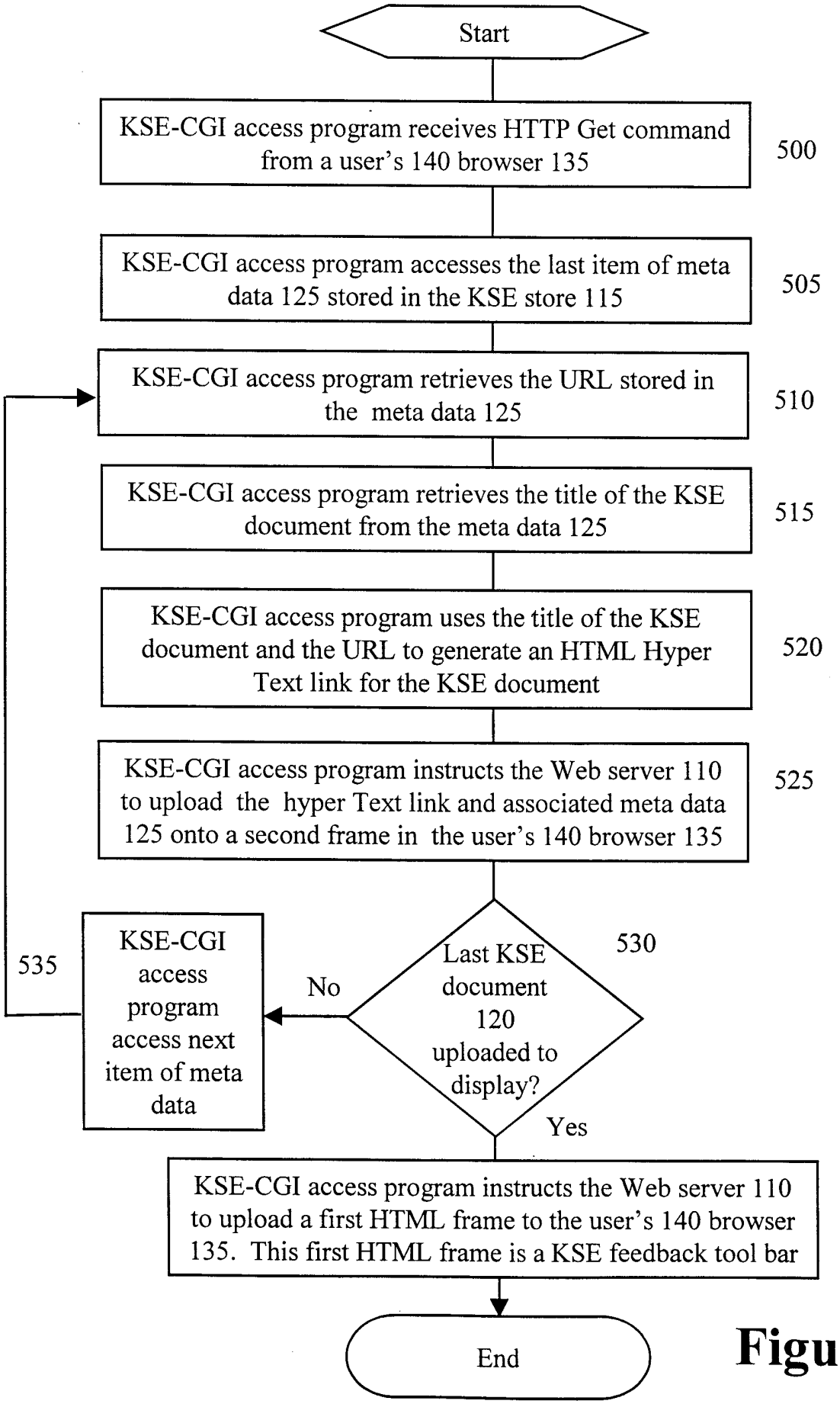

Start
KSE-CGI access program receives HTTP Get command from a user's 140 browser 135
500
KSE-CGI access program accesses the last item of meta data 125 stored in the KSE store 115
505
KSE-CGI access program retrieves the URL stored in the meta data 125
510
KSE-CGI access program retrieves the title of the KSE document from the meta data 125
515
KSE-CGI access program uses the title of the KSE document and the URL to generate an HTML Hyper Text link for the KSE document
520
KSE-CGI access program instructs the Web server 110 to upload the hyper Text link and associated meta data 125 onto a second frame in the user's 140 browser 135
525
Last KSE document 120 uploaded to display?
530
No
Yes
KSE-CGI access program access next item of meta data
535
KSE-CGI access program instructs the Web server 110 to upload a first HTML frame to the user's 140 browser 135. This first HTML frame is a KSE feedback tool bar
540
End

Figure 5

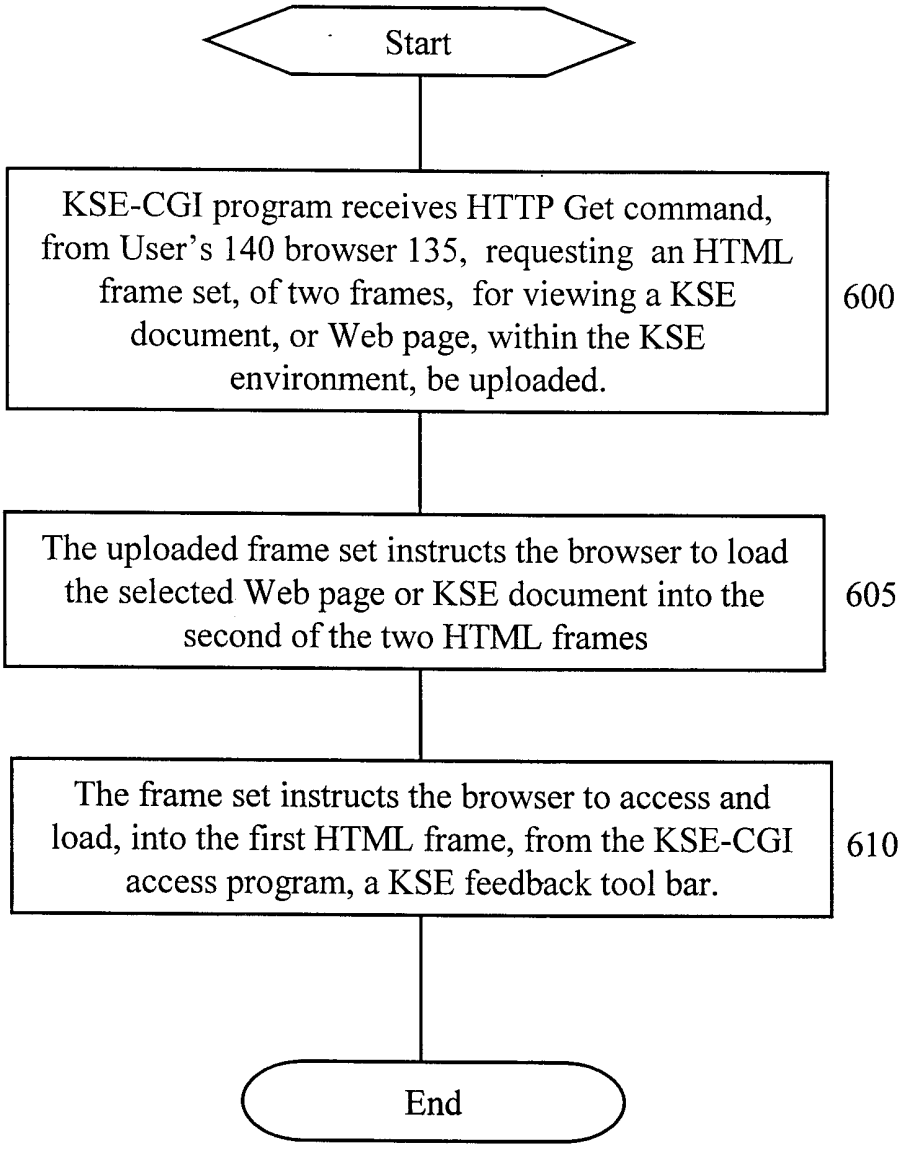

Start
KSE-CGI program receives HTTP Get command, from User's 140 browser 135, requesting an HTML frame set, of two frames, for viewing a KSE document, or Web page, within the KSE environment, be uploaded.
600
The uploaded frame set instructs the browser to load the selected Web page or KSE document into the second of the two HTML frames
605
The frame set instructs the browser to access and load, into the first HTML frame, from the KSE-CGI access program, a KSE feedback tool bar.
610
End

Figure 6

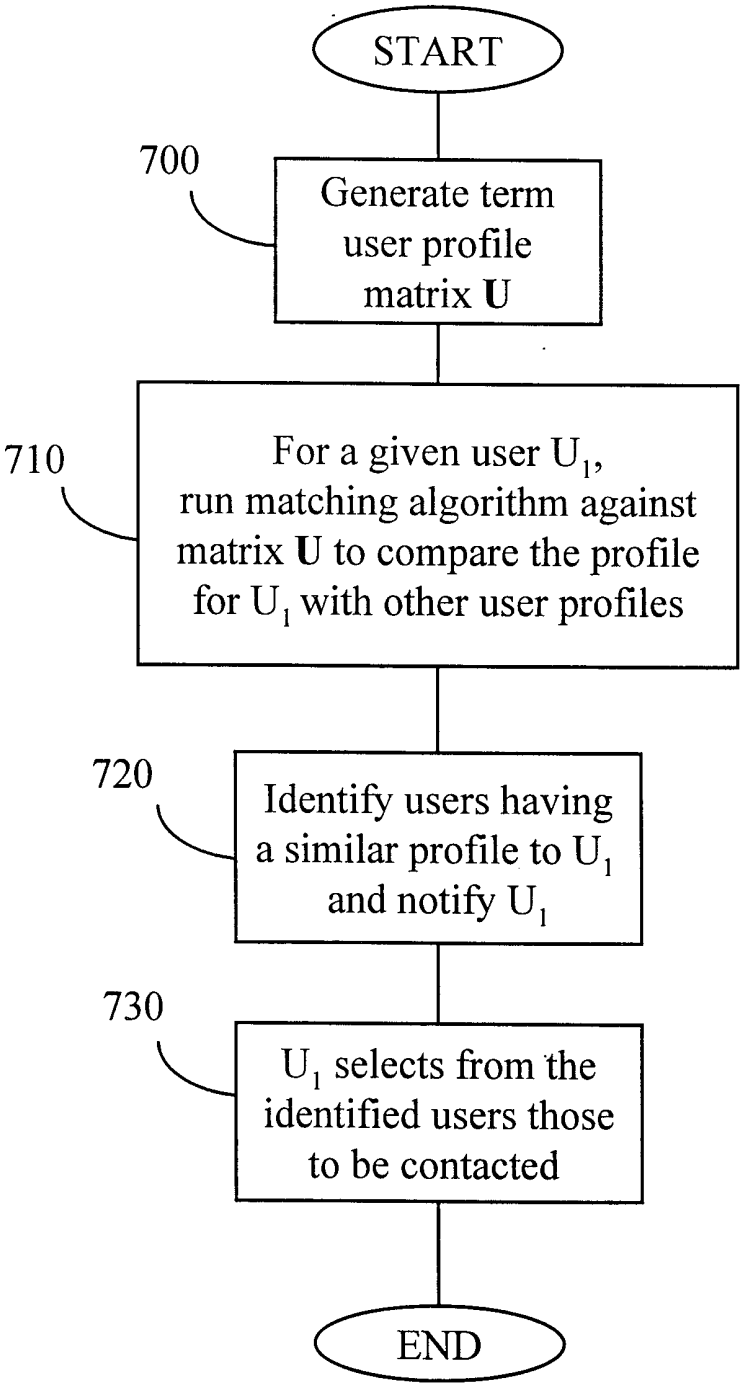

START
700
Generate term user profile matrix U
710
For a given user U_1, run matching algorithm against matrix U to compare the profile for U_1 with other user profiles
720
Identify users having a similar profile to U_1 and notify U_1
730
U_1 selects from the identified users those to be contacted
END

Figure 7

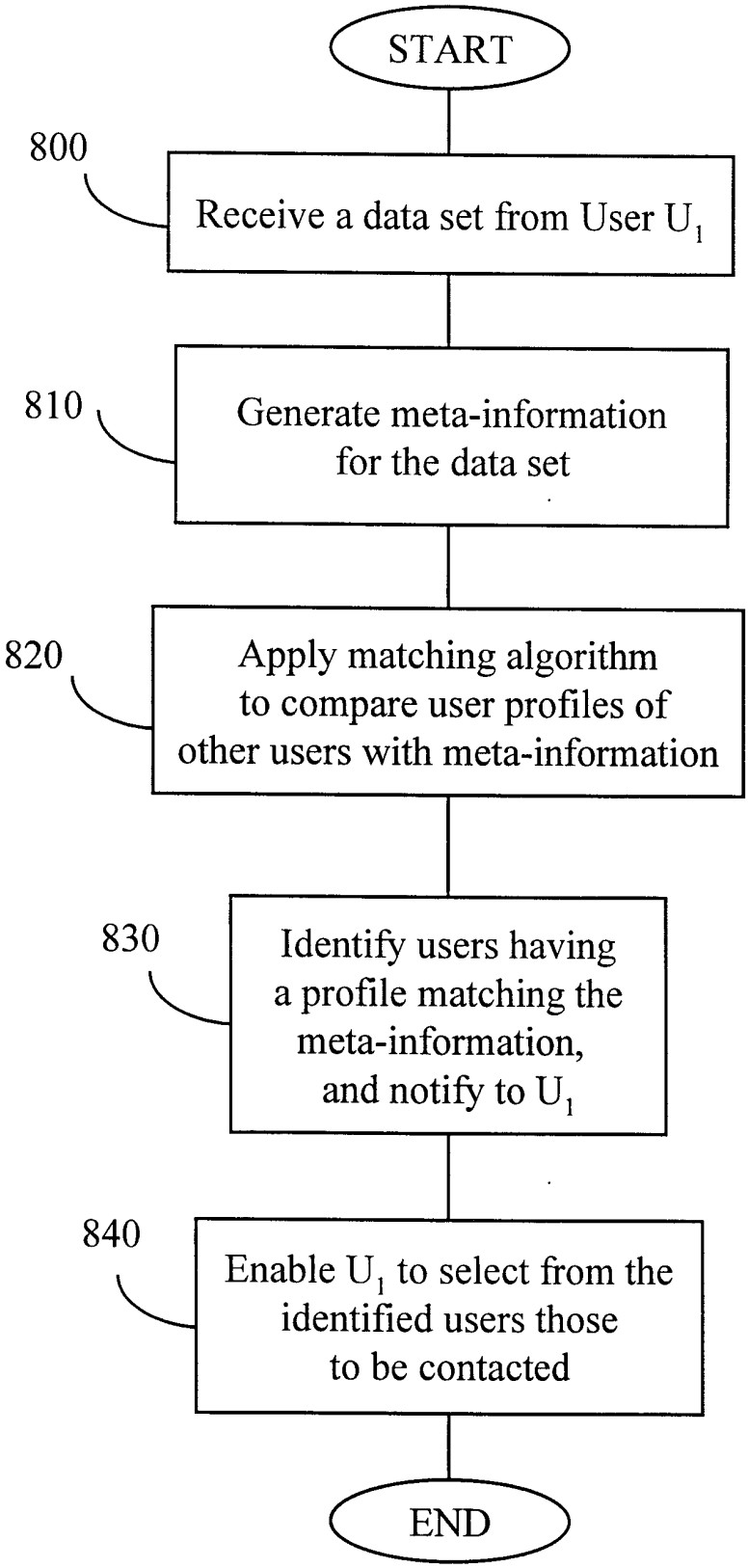

START
800
Receive a data set from User U_1
810
Generate meta-information for the data set
820
Apply matching algorithm to compare user profiles of other users with meta-information
830
Identify users having a profile matching the meta-information, and notify to U_1
840
Enable U_1 to select from the identified users those to be contacted
END

Figure 8

INTERNATIONAL SEARCH REPORT

Inte· ·tional Application No
PC1/GB 99/02495

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/30

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DAVIES N J ET AL: "KNOWLEDGE SHARING AGENTS OVER THE WORLD WIDE WEB" BT TECHNOLOGY JOURNAL, vol. 16, no. 3, July 1998 (1998-07), pages 104-109, XP000781604 the whole document ---	1-9
X	ADDISON E R: "Synthesizing summary knowledge from distributed heterogeneous information sources" MILCOM 97. MILCOM 97 PROCEEDINGS (CAT. NO.97CH36134), MILCOM 97 MILCOM 97 PROCEEDINGS, MONTEREY, CA, USA, 2-5 NOV. 1997, pages 1609-1613 vol.3, XP002096539 ISBN 0-7803-4249-6, 1997, New York, NY, USA, IEEE, USA the whole document -----	1-9

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

☐ Patent family members are listed in annex.

° Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 November 1999	09/11/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Katerbau, R

1