

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
31 July 2008 (31.07.2008)

PCT

(10) International Publication Number
WO 2008/090373 A1

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

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

(21) International Application Number:
PCT/GB2008/050036

(22) International Filing Date: 18 January 2008 (18.01.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0701158.8 22 January 2007 (22.01.2007) GB

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

(71) Applicant (for all designated States except US): **WEB-SENSE UK LIMITED** [GB/GB]; c/o Baker & Mackenzie, 100 New Bridge Road, London EC4V 6JA (GB).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **NASH, Alastair** [GB/GB]; SurfControl Limited, Riverside, Mountbatten Way, Congleton Cheshire CW12 1DY (GB).

(74) Agent: **ROBINSON, Ian**; Appleyard Lees, 15 Clare Road, Halifax Yorkshire HX1 2HY (GB).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(54) Title: RESOURCE ACCESS FILTERING SYSTEM AND DATABASE STRUCTURE FOR USE THEREWITH

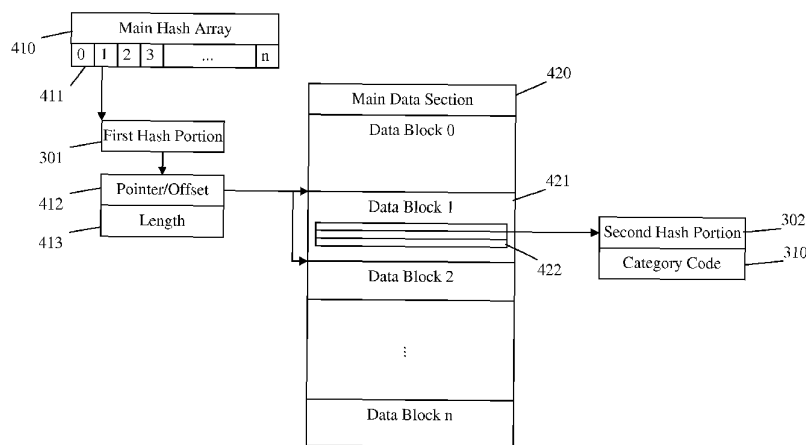


Fig. 7

(57) Abstract: A database structure (400) stores data representing a plurality of resource identifiers (200) such as URLs in a computing system. The URLs (200) are each associated with a respective category code (310). Each of resource identifiers (200) is hashed to provide a hash value (300) which is divided into at least first and second hash portions (301, 302). The first hash portions (301) index a main hash array (410) to select a main data block (421) in a main data section (420) holding a set of the second hash portions (302) and the respective category codes (310). For security and compactness, optionally the hash value (300) is further divided to provide a third hash portion (303) which is discarded and is not used to access the database (400). A database generation apparatus, a database search apparatus and a filtering apparatus, a method of creating a database structure and a method of retrieving data from a database structure, and to a computer readable storage medium having a database structure recorded thereon are each described.



WO 2008/090373 A1

RESOURCE ACCESS FILTERING SYSTEM AND DATABASE STRUCTURE FOR USE THEREWITH

Field of the Invention

The present invention relates in general to a resource access filtering system which
5 controls access to resources of a computing network with reference to a database structure
storing data representing a plurality of unique resource identifiers such as URLs. Also, the
present invention relates to a generator apparatus to create such a database structure, and to
a filter apparatus to control access to resources in a computing system using such a database
structure. Further, the present invention relates to a method of creating a database structure
10 and a method of retrieving data from a database structure, and to a computer readable storage
medium having a database structure recorded thereon.

Description of the Background Art

One of the great benefits of the Internet is that many millions of users have access to
the shared information and communication of the World Wide Web. However, open access to
15 all forms of information or communications is not always appropriate. For example, many
schools and businesses provide Internet access for their students and employees. The school
or business is, at least in part, responsible for dissemination of information within that
organisation and is usually under an obligation to prevent circulation of racist, sexist or other
abusive materials. This is just one example situation where there is a strong need for a
20 measure of control over Internet access. Another example is a home environment, where
parents may wish to prevent their children accessing adult-oriented web pages. Other
examples include public spaces such as libraries, Internet cafes or public Internet kiosks.

Several systems are available in the related art to monitor or control access to the
Internet. For example, US 5,884,033 (Duvall et al) discloses a client-based filtering system that
25 compares portions of incoming and/or outgoing messages to filtering information stored in a
database, and determines an appropriate action, such as whether to ALLOW or BLOCK
transmission of messages. An update server is accessible over the Internet for updating the
database.

US2003/0088577A1 (Thurnhofer et al) discloses a method of generating a database
30 having a relatively large number of data items classified into a relatively small number of
categories, using a form of directed graphs. This database was been developed specifically for
compactly and efficiently storing a large number of URLs and associated category codes.
However, the database must be completely rebuilt at each revision to add new URLs or amend
the categorisation.

The filtering systems and database structures of the related art encounter a variety of different demands. For example, such database structures are often required to be relatively compact so as to minimise a size of the database. However, it is also desired for the system to access the database structure at high speed. Further, it is desired to update the database at frequent intervals. Further still, it is desired to protect the security of the database. These and other problems in the related art are addressed by some of the exemplary embodiments of the invention, and will now be discussed in more detail.

Summary of the Invention

According to the present invention there is provided a resource access filtering system, a generator apparatus, a filter apparatus, a method of creating a database structure, a method of retrieving data from a database structure, and a computer readable storage medium having a database structure recorded thereon, as set forth in the appended claims. Preferred features of the invention will be apparent from the dependent claims, and the description which follows.

In one aspect of the present invention there is provided a resource access filtering system, comprising a generator unit, a client unit and a filter unit. The generator unit is arranged to create a database structure and comprises at least a categorisation unit arranged to provide a plurality of resource identifiers each associated with a respective category code; a hashing unit arranged to hash each of the plurality of resource identifiers to provide a hash value and to divide the hash value into at least first and second hash portions; and a storing unit arranged to output the database structure including at least a hash array and a main data section, wherein the main data section comprises the second hash portions and the respective category codes divided amongst a plurality of main data blocks, and the hash array comprises the first hash portions each associated with a pointer to a respective one of the main data blocks in the main data section. The client unit is arranged to request access to a resource with reference to a demanded resource identifier. Suitably, the resource is a resource within a computing network, and conveniently the resource is identified by a URL including a host portion and optionally other portions such as a page portion. The filter unit is arranged to control access to the resource with reference to a category code obtained from the database structure according to the demanded resource identifier, wherein the filtering unit comprises a communication unit arranged to obtain the demanded resource identifier in relation to a resource access request made by the client device; a database access unit arranged to return a category code appropriate to the demanded resource identifier from the database structure by generating a search hash value based on the resource identifier and dividing the search hash value into at least first and second hash portions, identifying a respective main data block in the main data section of the database structure by consulting the hash array using the first hash portion, and retrieving the category code from the respective main data block according to the second hash portion; and a control unit arranged to determine a control action in relation

to the resource access request made by the client device with reference to the category code obtained by the database access unit.

In one embodiment, the hashing unit is arranged to divide each of the hash values also into a third hash portion which is then discarded to improve security of the database structure.
5 The database access unit is likewise arranged to divide the demanded resource identifier into corresponding first, second and third hash portions where only the first and second hash portions are employed to access the database structure.

In one embodiment, the storing unit is further arranged to provide a page data section and/or a conflict data section to enhance the accuracy and/or granularity of the database
10 structure.

In one embodiment, the filter unit comprises at least a faster-access primary storage device and a slower-access secondary storage device operates in in a first memory-based mode using solely the primary storage device for faster access to the database structure, or operates in a second shared-access mode where the primary storage device is used in
15 cooperation with the secondary storage thereby reducing memory usage in the primary storage device whilst maintaining acceptable real-time access speed to the database structure.

In one embodiment, the system further comprises an updater unit arranged to produce an update structure to update the database structure. The update structure allows an
20 incremental update of the database structure and exemplary embodiments are arranged to update the database structure whilst the database structure is still live in use.

Other aspects of the present invention relate particularly to the generator unit, the filter unit and the updater unit as separate elements of the system described herein.

A further aspect of the present invention provides a computer readable storage medium
25 having recorded thereon a database structure storing data representing a plurality of resource identifiers each associated with a respective category code, comprising: a main data section having a plurality of main data blocks, wherein each of the plurality of main data blocks comprises a plurality of second hash portions and each of the plurality of second hash portions is associated with a respective category code; and a main hash array comprising a plurality of
30 first hash portions, wherein each of the first hash portions is associated with a reference pointer to a respective one of the plurality of main data blocks in the main data section; wherein each of the first hash portions and the second hash portions is divided from a hash value derived a respective one of the plurality of resource identifiers.

A further aspect of the present invention provides a method of creating a database structure, comprising the computer-implemented steps of: providing a plurality of resource identifiers each associated with a respective category code; hashing each of the plurality of resource identifiers to provide a respective hash value and dividing the respective hash value
5 into at least a first hash portion and a second hash portion; and storing at least a main hash array and a main data section in the database structure, wherein the main data section comprises a plurality of main data blocks each comprising a set of the second hash portions associated with the respective category codes, and the main hash array comprises the first hash portions each associated with a respective main data block in the data section.

10 A further aspect of the present invention provides a method of retrieving data from a database structure, comprising the computer-implemented steps of: receiving a demanded resource identifier which identifies a resource in a computing system; providing a hash value based on the resource identifier and dividing the hash value into at least first and second hash portions; and identifying a main data block in a main data section of the database structure by
15 consulting a hash array using the first hash portion, and retrieving a category code from the identified main data block according to the second hash portion.

At least some embodiments of the invention may be constructed solely using dedicated hardware, and terms such as 'module' or 'unit' used herein may include, but are not limited to, a hardware device, such as a Field Programmable Gate Array (FPGA) or Application Specific
20 Integrated Circuit (ASIC), which performs certain tasks. Alternatively, elements of the invention may be configured to reside on an addressable storage medium and be configured to execute on one or more processors. Thus, functional elements of the invention may, in some embodiments include, by way of example, components, such as software components, object-oriented software components, class components and task components, processes,
25 functions, attributes, procedures, subroutines, segments of program code, drivers, firmware, microcode, circuitry, data, databases, data structures, tables, arrays, and variables. Further, although the preferred embodiments have been described with reference to the components, modules and units discussed below, such functional elements in alternate embodiments may be combined into fewer elements or separated into additional elements.

30 **Description of the Drawings**

For a better understanding of the invention, and to show how embodiments of the same may be carried into effect, reference will now be made, by way of example, to the accompanying drawings in which:

Figure 1 is a schematic overview of a system and apparatus as employed in a first
35 example embodiment of the present invention;

Figure 2 is a schematic overview of a system and apparatus as employed in a second example embodiment of the present invention;

Figure 3 is a schematic overview of a filter unit as employed in example embodiments of the invention;

5 Figure 4 is a schematic overview of a generator apparatus as employed in example embodiments of the invention;

Figure 5 shows a series of data operations which are performed by the generator apparatus of Figure 4;

10 Figure 6 is a schematic overview of an exemplary database structure as employed in embodiments of the invention;

Figure 7 shows the exemplary database structure in more detail concerning a main hash array and a main data section;

Figure 8 is a schematic overview of a method of generating the database structure;

15 Figure 9 shows a preferred mechanism for extracting a category code from the database structure appropriate to a demanded URL;

Figure 10 shows the exemplary database structure in more detail concerning a conflict data section;

Figure 11 shows the exemplary database structure in more detail concerning a page data section;

20 Figure 12 shows the filter unit in more detail concerning a memory access mode and a shared access mode;

Figure 13 is a schematic overview of a system and apparatus as employed to update the exemplary database structure;

25 Figure 14 is a schematic overview of an exemplary update database structure as employed in embodiments of the invention;

Figure 15 shows the exemplary update database structure in more detail concerning a main update section;

Figure 16 shows the exemplary update database structure in more detail concerning a page update section; and

Figure 17 shows the exemplary update database structure in more detail concerning a conflict update section.

Referring to Figure 1, a system and apparatus as employed in a first example embodiment of the present invention is applicable such as in a small-scale corporate network or a domestic home environment. In this first example embodiment, one or more client machines 10 are connected to the Internet 20 through a client gateway 12 and an Internet Service Provider (ISP) 21. Each client machine 10 provides input and output interface functions as appropriate for a human user, suitably including a display screen providing a graphical user interface (GUI), and a user input unit such as control keys, a mouse or other suitable devices. As shown in Figure 1, in one embodiment the client machine 10 is a computing platform such as a desktop computer, a laptop computer, or a personal digital assistant (PDA). In another embodiment, the client machine 10 is a function-specific Internet appliance, such as a web-TV. In a third example, the client machine 10 is a public Internet kiosk. Many other specific forms of client machine 10 are also applicable to embodiments of the present invention.

Each of the client machines 10 provides a client application which uses URLs to obtain resources. In this example, a content server 30 is accessible via the Internet 20 and provides a resource such as a web page 32. Suitably, the client machine 10 provides a web browser application and makes a HTTP request to obtain the web page 32 from the content server 30. In this example, the web page 32 comprises text and/or graphics data in an appropriate format such as HTML, SGML, PostScript or RTF, amongst others. It will be appreciated however that the present invention is also applicable to many other resources provided in a computer system or computer network using URLs, such as file transfers under FTP, connection to a TELNET server, or Internet Relay Chat (IRC), amongst others.

The client gateway 12 suitably includes a modem or router, such as an analogue, ISDN, ADSL or SDSL modem, which connects to the ISP 21 over the plain old telephone system (POTS) or other wired or optical network to provide a network layer connection to the Internet 20. As another example, the client gateway 12 connects to the Internet 20 through a wireless network or cellular mobile network such as GSM or GPRS. In still other embodiments, the client gateway 12 connects to the Internet 20 through an intermediary such as a LAN or WAN, or optionally over a virtual private network (VPN). In one embodiment, the client machine 10 and the client gateway 12 are formed as physically separate devices and communicate by any appropriate wired or wireless link. In other embodiments, the client gateway 12 is integrated within the client machine 10.

As shown in Figure 1, a filter unit 40 is arranged to hold a database structure 400 which stores URLs and associated category codes. Conveniently, the filter unit 40 is integrated within

the client gateway 12 or within one or more of the client machines 10. Figure 1 also shows a database generator apparatus and database updater apparatus 50 which generates and updates the database structure 400. The database structure 400 is suitably provided to the filter unit 40 from a portable data storage medium such as an optical disk, or by downloading
5 from a communication network such as over the Internet 20.

Figure 2 is a schematic overview of a system and apparatus as employed in a second example embodiment of the present invention. The example of Figure 2 is typical of a computing network as used in commercial offices and other large organisations. This second example provides one or more client machines 10 coupled through a local area network (LAN)
10 11. Typically, the LAN 11 includes one or more proxy servers 14 coupled to the Internet 20 such as through an ISP (not shown). Resources are accessed using URLs, either locally within the LAN or from outside the LAN such as over the Internet 20. A server 16 coupled to the LAN 11 provides the filter unit 40 and holds the database structure 400. As for Figure 1 above, the database structure 400 is generated by the database generator apparatus 50 and
15 is provided to the filter unit 40 such as over the Internet 20. The generator apparatus also acts as a database updater apparatus 50 which prepares an update package that is sent to the filter unit to update the database structure 400. There are many specific variations on the example network typography shown in Figure 2, as will be familiar to those skilled in the art. For example, in a further embodiment, the filter unit 40, the database updater apparatus and
20 the database structure 400 are located in the Internet cloud 20 upstream of the LAN 11 and are accessed as an on-demand managed service remote from the LAN 11.

In the example embodiments of Figures 1 and 2, the filter unit 40 is arranged to perform one of a variety of filtering functions, which may include passively monitoring and logging URLs for later inspection, analysis or reporting, or performing an active controlling function
25 such as determining whether the client machine 10 will be allowed to access the requested resource. The filter unit 40 is arranged to perform such filtering functions by placing the URLs into categories, by consulting the database structure 400.

Figure 3 is a more detailed view of the filter unit 40. The filter unit 40 comprises a communication unit 41, a database access unit 42, and a control unit 43. The communication
30 unit 41 obtains a URL in relation to a resource access request, such as by intercepting a HTTP request made by the client machine 10 to access the web page 32 on the resource server 30. Alternatively, the resource access request is intercepted elsewhere in the system and a request is made to the filter unit 40 to provide a category code for the demanded URL. The database access unit 42 interrogates the database structure 400 to determine a category of the URL. Optionally, the control unit 43 performs a filter operation with reference to the URL
35 and the determined category code, such as an active controlling function or a passive logging function. The communication unit 41 provides an appropriate output, such as sending a HTTP

reply to the client machine 10 denying access to the requested URL, or forwarding an allowable resource access request towards the resource server 30, or returning a category code relevant to the demanded URL.

In a simple example, the filter unit 40 places the URLs into one of two categories, such as either "allow" or "deny". In a more sophisticated embodiment, the URLs are categorised with greater granularity. In the present example, the database structure 400 places each URL into one of about forty categories. The example embodiment provides eight core categories such as "adult/sexual explicit", "criminal skills", "drugs, alcohol, tobacco", "violence" or "weapons", as well as thirty two productivity-related categories such as "advertisements", "games", "hobbies and recreation" or "kid's sites". Providing this predetermined set of categories allows a more sophisticated rules-based controlling or logging function to be performed. For example, a rule is used to alert an administrator when a request is made for any of the core categories, or to block selected productivity categories at particular times and allowing access only say at lunchtimes or outside work hours. To cater for all eventualities, the preferred categories may also include "don't know" or "not found" options. The specific downstream filtering actions for filtering resource access requests, such as passive monitoring functions or active blocking functions, may take any suitable form.

Figure 4 is a schematic overview of the generator apparatus 50 which generates the database structure 400. A categorisation unit 51 provides a list of URLs and associated category codes. A hashing unit 52 hashes a given URL to provide a URL hash. A storing unit 53 stores at least part of the URL hash into the database 400 along with the category code for that URL.

Figure 5 shows a series of data operations which are performed by the generator apparatus 50 when generating the database structure 400.

Firstly, Figure 5 shows a resource identifier 200 such as a Uniform Resource Identifier (URI) as defined in RFC3986. In general terms, the URI 200 is a numeric string or more commonly an alphanumeric string which is used to identify a resource within a network. The described invention is applicable in many specific environments and is applicable to many different specific forms of resource identifiers. For explanation and appreciation of the basic principles of the present invention, the example embodiments described herein are concerned with access to resources on a computer network. Most currently available computer networks are Internet Protocol (IP) based networks which employ Uniform Resource Locators (URLs) to identify resources. URLs are well known and will be familiar to the skilled person, such as from RFC1738. In Figure 5, the example URL 200 includes a host portion 202 and a page portion 204. The host portion 202 identifies a particular host (e.g. "www.host.com"), whilst the page portion 204 identifies a relative path to a specific page (e.g. "/directory/page.html"). As another

specific example, the host portion 202 may be expressed as an IPv4 address, which typically takes a four-octet dotted decimal format. As another example, the host portion 202 is an IPv6 address.

5 Secondly, Figure 5 shows a URL hash value 300 which is produced by hashing the resource identifier, in this case the URL 200, with a hashing algorithm. The example
embodiments employ a commonly available hashing scheme (also termed a message digest
algorithm) such as MD4 or MD5. The example embodiments discussed herein use MD4,
which is typically some 12% faster than MD5. Although less secure than MD5, the database
structure discussed herein still allows the faster MD4 to be used securely. In the illustrated
10 example of Figure 4, an alphanumeric string of arbitrary length representing a URL host
portion 202 is hashed using the MD4 hashing algorithm to provide a fixed-length 16-byte URL
hash value 300. The length of the hash value 300 is suitably chosen according to the number
of records in the database structure 400. In a larger database, the hash length is extended to,
for example, 24 or 32 bytes, such that the hash 300 has a minimal probability of producing
15 identical hash values 300 (a hash collision) for two different URL strings 200.

Thirdly, Figure 5 shows the URL hash 300 divided into at least first and second portions
301, 302, and optionally also into a third portion 303. The first portion 301 is conveniently the
least significant byte or bytes of the hash 300, such as the lowest two bytes 0 and 1 as shown
in Figure 5. The second portion 302 is conveniently a group of next most significant bytes,
20 which in this case is the six bytes labelled 2 to 7. The third portion 303 is conveniently a group
of most significant bytes, which in this case are the bytes 8 to 15. Many other divisions of the
URL hash are also possible. The first and second portions 301, 302 of the URL hash 300 are
retained to be applied to the database structure 400 as will be described in more detail below,
whilst the third portion 303 of the URL hash 300 is discarded.

25 Figure 6 is a schematic structural overview of the database structure 400 according to
an example embodiment of the present invention. The database structure 400 includes a main
hash array 410 and a main data section 420, a conflict data section 430, a page data section
440, a category data section 450 and a file header 460.

Figure 7 shows the exemplary database structure 400 in more detail, concerning the
30 main hash array 410 and the main data section 420. The other parts of the database will be
described later.

The main hash array 410 is a hash table used to index the database structure, having
entries 0 to n consistent with the first hash portion 301 of the URL hash value 300. For a 2-
byte index, n is 65,536 entries. Each entry 411 in the hash table 410 identifies a block of data
35 in the main data section 420, suitably with a pointer 412. A length value 413 gives the length of

the data block. In the example embodiments, the data blocks are not all of the same size. The pointer 412 denotes an offset value (in bytes) from a start of the main data section 420. The offset 412 together with the length value 413 select a block of data 421 from the main data section 420.

5 Each block of data 421 in the main data section 420 stores a plurality of records 422. Each record 422 comprises a second hash portion 302 of a URL hash value 300 corresponding to a particular host portion 202 of a URL 200. Also, each record 422 comprises a corresponding category code 310 for the relevant URL. In this case, the category code 310 is a root category code which applies to the root or host portion 202 of the relevant URL 200.

10 Figure 8 is a schematic overview of a method of generating the database structure 400.

A URL 200 is hashed at step 801. At step 802, the URL hash is divided into at least two parts. The first part of the hash is stored in the database structure. As discussed above, a first portion 301 relates to the main hash array 410, and the second portion is stored in the relevant data block 421 of the main data section 420, at steps 803 and 804. The second part of the hash, namely the third portion 303, is discarded at step 805.

15 Figure 9 shows a preferred mechanism for extracting a category code from the database structure appropriate to a demanded URL.

The demanded URL 200 is hashed at step 901 to provide a URL hash value 300. In particular, the host portion 202 of the URL 200 is hashed to provide the fixed-length 16-byte hash value 300. At step 902 the URL hash value is divided into the first, second and third portions 301, 302, 303. At step 903, the first URL hash portion 301, in this case the least significant two bytes, is used to index the main hash array 410, which suitably retrieves an appropriate offset 412 and length value 413. At step 904, a data block 421 is selected from the main data array 420, using the offset 412 and length 413 values. The selected data block 20 421 is searched at step 905 using the second hash part 802. Conveniently, the data block is numerically ordered, and the search is a binary chop or similar technique. A binary chop is a fast and simple mechanism for searching a numerically order list. Each stored second hash portion 302 is a numerical value of a known size, in this case 6 bytes. A simple integer comparison is used to search the hash values, which is much faster than an alpha-numeric string to compare full URLs 200 or URL host portions 200.

25 At step 906, a successful match of the second hash portion 302 in the searched data block 421 selects a particular record 422 and allows the corresponding category code 310 to be retrieved. The retrieved category code 310 is the category code appropriate to the demanded URL 200. Otherwise, a "not found" condition is returned.

Figure 10 shows further details of the database structure 400 relating to the conflict data section 430.

Even with the preferred 16-byte MD4 hashing algorithm, there is a small possibility that two different URLs 200 will hash to the same value. Also, since the third hash portion 303 is discarded, there is a possibility that two different URLs 200 will hash to the same partial value 301, 302 (i.e. have the same lower 8 bytes). In practice, such coincidences are very rare. However, in the event that a coincidence is detected when storing a second URL which hashes to the same value as a previously stored first URL, then the conflicted data section 430 is employed.

As shown in Figure 10, the appropriate entry 422 in the main data section 420 is marked as invalid using a marker 425, such as by writing a predetermined value in place of the usual category code 310. The marker 425 shows that the conflict data section 430 is to be consulted.

The conflict data section 430 stores full hash values 300 for the first and second URLs. In a this example embodiment, each of the full hash values is determined under a superior hashing scheme, in this case MD5, in order to further reduce the possibility of conflicts in the full hash values. The conflict data section also stores the category code 310 for each of these URLs.

In the present scheme, it is possible that a second URL which has not been encountered previously will conflict with a stored first URL, resulting in an inaccurate category 310 being returned for the second URL. When this occurs, the database structure 400 is updated to then include the second URL, by moving both the first URL and the second URL to the conflict list 430.

Figure 11 shows further details of the example database structure 400 concerning the page data section 440.

The page data section 440 contains category codes allocated to individual pages 204 dependent from a host 202. A root category code is provided for each host 202, and optionally page category codes are provided for some or all of the pages 204 associated with a particular host 202. For example, the host domain "www.bbc.co.uk" is allocated category, e.g., "37", whilst a particular page of the domain at "www.bbc.co.uk/sport.html" is allocated another code, e.g., "38".

Referring to Figure 11, each entry 422 in the main data section 420 further includes a page index 426 which indexes a block of page data 444 in the page data section 404. A

default value (e.g. "0") is used to show that there are no pages in the page data section for a particular record 422.

In the page data section 404, the pages are grouped by index and each held as page strings 441 (e.g., "/sport", "/weather", and so on). In this example, each page string 411 has an associated page category code 311. The page strings are sorted alpha-numerically to enable a sequential search of the page data block 444. Each page block may be of any length and suitably ends with a delimiting end character (e.g. null).

Referring again to Figure 6, the category data section 450 contains a list of category codes and related textual category names. The category data section 450 is useful in interpreting the category codes stored in the database structure. Suitably, the category data section 450 provides category names in a plurality of different languages, so that category names are displayed according to local preferences. The file header section 460 contains global information for the database structure file such as a file name. Suitably, the file header 460 identifies locations of the other data sections 410-450 in the file. Further, the file header contains validation data, such as checksums.

Figure 12 shows further details of the filter unit 40 of Figure 3 in relation to the example database structure 400 as discussed above. As shown in Figure 12, the filter unit 40 is arranged to access at least first data storage unit 44 and a second data storage unit 45, which are able to hold the database structure 400 and provide data to the database access unit 42. Here, the data storage units 44, 45 include at least one form of fast-access primary storage device 44 such as a random-access solid-state memory, and at least one form of slow-access large-capacity long-term secondary storage device 45, such as a magnetic disk drive or optical disk storage.

In a first memory-based mode, the entire database structure 400 is loaded into the faster-access storage of the memory 44. Suitably, the database structure 400 is initialised in the memory 44 by converting each of the offsets 412 in the main hash array 410 into memory location pointers which directly or indirectly reference memory address locations where the data blocks 421 of the main data section 420 are now held. Holding the database structure entirely in the working memory 44 maximises lookup speed of the database structure 400. However, even though the database structure 400 is relatively compact, there is still a penalty in terms of the amount of memory required. In many practical situations, such as for devices with a limited memory capacity, it is desired to minimise memory area whilst maintaining adequate access times to retrieve data from the database structure. For example, equipment such as a Public Internet Kiosk or a hand held portable device (e.g. a PDA or a mobile phone) typically has relatively little free memory.

As also shown in Figure 12, in a second shared-access mode only selected parts of the database structure 400 are loaded into the memory 44. In this shared-access mode, the main hash array 410 is loaded into the memory 44, whilst the main data section 420, the page data section 440, and the conflict data section 430 are all retained on the secondary storage 45.

5 At initialisation of the database structure, the offsets 412 in the main hash array 410 are converted and stored in the memory 44 as offsets from the start of the database structure file 400. The data blocks 421 are not initially loaded into the memory 44. Instead, when a lookup is performed, the converted offset 412 is obtained from the main hash array 410 in memory 44 and used to load only the required data block 421 from the main data section 420. Once the
10 lookup is completed, the loaded data is then freed. Similarly, where the page data section 430 is employed, a page level access to the page index table 432 loads a required page data block 444 when needed. In a preferred embodiment, the page indexes 432 of the page data section 430 are also loaded into the memory 44.

 This shared-access mode significantly reduces memory footprint. Further, the database
15 structure is arranged to optimise disk-based retrieval. The offsets 412 are conveniently relative to the start of the file and placing the main data section 420 at an early part of the file helps to reduce disk seek times. Further, the main data section 420 and the page data section 440 are closely located to reduce disk head movement.

 The described database structure to store URLs and associated category codes is, by
20 its nature, a relatively large set of data. For example, the database structure may be required to hold many millions of URLs. As a practical example, the database structure contains of the order of 5.5 million main hosts 202 or IP addresses. Also, the page data section 440 lists of the order of 500,000 pages. However, example embodiments of the database structure require only about 65Mb of storage. Hence, the described database structure 400 is less costly
25 to transmit across a computer network, and is less costly to store.

 The preferred database structure is compact, whilst maintaining speedy access. In the
example memory-mode discussed above, the database structure is accessible at the order of
70,000 accesses per second with domain level categorisations. In the shared-mode, of the
order of just 2Mb of data is held in memory, with the remainder (i.e. around 63Mb) remaining
30 on secondary storage such as a hard disk. Only three disk read operations are required per categorisation, in order to return a category code. Most categorisations take only one or two disk reads. Root level categorisations (domain level) take only one disk read. The shared-mode in an example embodiment handles over 25,000 categorisations per second.

In many cases, the database structure needs to be updated frequently, such as in order to add new URLs and category codes, or to amend the existing data. Typically, updates are required at least once per day.

5 Figure 13 shows an example system to transmit a database update structure 500, which is employed to incrementally update the database structure 400.

The generator unit 50 includes an updater unit 54 which compares a current version of the database structure 400a against a required new version 400b to produce the update structure 500. The update structure 500 is then transmitted to the filter unit 40 over a communication network such as the Internet 20. Sending just the incremental update 500
10 significantly reduces bandwidth consumption in the communication network 20. Optionally, the update structure 500 is compressed to further compact its size during transmission.

The filter unit 40 applies the update structure 500 to a current copy of the database structure 400a to produce the required new database structure 400b incorporating the desired updates, as will now be explained in more detail.

15 Figure 14 shows a general overview of the update structure 500, including a file header 510, a header update section 560, a main update section 520, a page update section 540, a conflict update section 530 and a category update section 550.

The file header 510 contains header information relating to the update file 500 such as date, version, file size and file checksums.

20 The header update section 560 contains data to completely replace the header section 460 of the current database structure 400a. The header 460 contains offsets and sizes of the other sections in the structure (e.g. offset and size of the main data section 420). Sending the entire new header 460 reduces workload at the filter unit 40 by avoiding recalculation of the section offsets and sizes after the update has completed.

25 The category update section 550 contains a complete replacement of the category data section 460. This data changes relatively infrequently, and is retransmitted in its entirety.

Figure 15 shows the main update section 520 in more detail. The main update section 520 contains a list of update commands 521 each followed by one or more parameters 522. The commands are suitably expressed as a predetermined code (e.g. 0x01, 0x02 and 0x03)
30 and are used to insert, modify or delete a record 422 in the main data section 420. The commands are sorted such that the order of the commands 521 corresponds to the order of the records 422, i.e. according to the hash portions 301, 302.

An INSERT command instructs the filter unit 40 to insert a new record 422 in the main data section 420 of database structure, at a location identified by the POSITION parameter 522. The POSITION parameter 522 gives an offset position relative to a position of a previous command (or a default start position) where the instructed command is to be performed.

5 Hence, the sequential ordering of the commands 521 allows the filter unit 40 to step through the database structure and make appropriate changes in order. The INSERT command also provides parameters 522 for second hash portion 302, the category code (CC) 310 and, if required, the page level index (PLI) 315 to complete the new record 422. The first hash portion 301 is also provided, and the main hash array 410 is recalculated by the filter unit 40 to allow

10 for the new record 422.

A MODIFY command instructs the filter unit 40 to modify an existing record 422 in the main data section 420. In this case, the parameters 522 provide new data to be written into the database structure to replace the current data of the relevant record 422. In this example, the parameters 522 provide a new category code (CC) 310 and a new page level index (PLI) 315.

15 A DELETE command instructs the filter unit 40 to delete a record 422 from the main data section 420 of database structure, as identified by the POSITION parameter 522. The first hash portion 301 is also given, such that the filter unit 40 recalculates the main hash array 410 to account for the now deleted record 422.

Figure 16 shows more detail of the page update section 540 which supports INSERT, 20 MODIFY and DELETE commands 541 with appropriate parameters 542, similar to the main update section 520 discussed above. The page commands 541 are sorted by the first hash portion 301 and then by the page string 441.

A page INSERT command instructs the filter unit 40 to insert a new page block 444 in the page data section 440 of the database structure, or insert a new page string 441 and page 25 category code 331 in an already existing page block 444. A first parameter 522 gives a Page Level Index (PLI) 426 when adding an entire new page block 444, or else "0" to add a page entry. A second LGTH parameter gives the length of a following third parameter (e.g. in bytes). The third parameter provides a set of page strings 441 and page category codes 311 for a new page block 434, or else gives the page string 441 and page category code 311 for the inserted 30 page. The POSITION parameter 522 gives an offset position relative to a position of a previous command (or a default start position) where the instructed command is to be performed.

A page MODIFY command 521 instructs the filter unit 40 to modify a page block size with a SZE parameter 522, or else modify a page category code (PCC) 311 of an existing 35 page in the page data section 440.

A page DELETE command instructs the filter unit 40 to delete a particular page string 441 and related page category code 311 from the page data section 440, as identified by the POSITION parameter 522. Alternatively, a Page Level Index 426 is given in the first parameter 522 and the DELETE command the deletes an entire page block 444.

- 5 Figure 17 shows more detail of the conflict update section 530 which supports conflict INSERT, MODIFY and DELETE commands 531 with appropriate parameters 532, similar to the main update section 520 and the page update section 540 discussed above. The conflict commands 531 are sorted by the full hash value 300.

10 A conflict INSERT command instructs the filter unit 40 to insert a new full hash value 300 in the conflict data section 430 with the category code 300. The page level index 426 is provided where the relevant root URL links to page level data. The POSITION parameter 522 gives an offset position relative to a position of a previous command (or a default start position) where the instructed command is to be performed. A conflict MODIFY command 521 instructs the filter unit 40 to modify a category code 310 or page level index 426 at the identified
15 position. A conflict DELETE command instructs the filter unit 40 to delete a particular entry in the conflict data section 430 and gives the first hash value 301 to update the main hash array 410.

 Referring again to Figure 13, the update structure 500 is applied to the existing
20 database 400a to produce the updated database 400b. To minimise downtime while the database structure 400 is being updated, the update structure 500 allows the database to be updated in small chunks as they are each copied from the existing structure 400a to the revised structure 400b, in particular due to the ordering of the commands. The main hash array 410 is recalculated during updating, to reference the new locations within the structure. Conveniently, comparing a checksum value of the new file 400b against a checksum given in
25 the update header 510 confirms that the update has been performed correctly. Once the update is complete, the new file 400b is suitably written over the old file 400a.

 In most cases, categorisation of URLs using the database structure may continue even during updating. However, in devices with limited memory, categorisation using the database structure is suitably halted completely for a short while, to allow the memory to be used for
30 updating each of the chunks of the structure in turn.

 The described database structure has many advantages, as discussed above or as will be appreciated by the skilled person by implementing the example embodiments of the invention.

 In some of the exemplary embodiments, security of the database structure is improved
35 by being difficult to reverse engineer. Compiling data into the database takes a high degree of

skill and effort and it is desirable that the database should be relatively difficult to reverse engineer, whilst still being easy to use legitimately. By sorting only part of the URL hash 301, 302, it is impossible for an original URL 200 to be obtained from the stored data.

5 The database structure 400 discussed herein is relatively compact, which minimises storage space in temporary memory (e.g. RAM) or on a permanent storage medium (e.g. a hard disk). Further, the database structure provides a fast look-up, so that data relevant to a particular record can be quickly retrieved from the database. Further still, embodiments of the database structure discussed herein are readily updated, such as by providing an incremental daily update to keep the database current. Yet further, the example structure may be updated
10 easily and whilst still live online.

Although a few preferred embodiments have been shown and described, it will be appreciated by those skilled in the art that various changes and modifications might be made without departing from the scope of the invention, as defined in the appended claims.

15 Attention is directed to all papers and documents which are filed concurrently with or previous to this specification in connection with this application and which are open to public inspection with this specification, and the contents of all such papers and documents are incorporated herein by reference.

20 All of the features disclosed in this specification (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, may be combined in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

25 Each feature disclosed in this specification (including any accompanying claims, abstract and drawings) may be replaced by alternative features serving the same, equivalent or similar purpose, unless expressly stated otherwise. Thus, unless expressly stated otherwise, each feature disclosed is one example only of a generic series of equivalent or similar features.

30 The invention is not restricted to the details of the foregoing embodiment(s). The invention extends to any novel one, or any novel combination, of the features disclosed in this specification (including any accompanying claims, abstract and drawings), or to any novel one, or any novel combination, of the steps of any method or process so disclosed.

CLAIMS

1. A resource access filtering system, comprising:
 - a generator unit (50) arranged to create a database structure (400), wherein the generator unit (50) comprises:
 - 5 a categorisation unit (51) arranged to provide a plurality of resource identifiers (200) each associated with a respective category code (310);
 - a hashing unit (52) arranged to hash each of the plurality of resource identifiers (200) to provide a hash value (300), and to divide the hash value (300) into at least first and second hash portions (301, 302); and
 - 10 a storing unit (53) arranged to output the database structure (400) including at least a hash array (410) and a main data section (420), wherein the main data section (420) comprises the second hash portions (302) and the respective category codes (310) divided amongst a plurality of main data blocks (421), and the hash array (410) comprises the first hash portions (301) each associated with a pointer (412) to a respective one of the main data
 - 15 blocks (421) in the main data section (420);
 - a client unit (10) arranged to request access to a resource (32) with reference to a demanded resource identifier (200); and
 - a filter unit (40) arranged to control access to the resource (32) with reference to a category code (310) obtained from the database structure (400) according to the demanded
 - 20 resource identifier (200), wherein the filtering unit (40) comprises:
 - a communication unit (41) arranged to obtain the demanded resource identifier (200) in relation to a resource access request made by the client device (10);
 - a database access unit (42) arranged to return a category code (310) appropriate to the demanded resource identifier (200) from the database structure (400), by generating a search
 - 25 hash value (300) based on the resource identifier (200) and dividing the search hash value (300) into at least first and second hash portions (301, 302), identifying a respective main data block (421) in the main data section (420) of the database structure (400) by consulting the hash array (410) using the first hash portion (301), and retrieving the category code (310) from the respective main data block (421) according to the second hash portion (302); and
 - 30 a control unit (43) arranged to determine a control action in relation to the resource access request made by the client device (10) with reference to the category code (310) obtained by the database access unit (42).

2. The resource access filtering system of claim 1, wherein:
- the hashing unit (52) is arranged to divide each of the hash values (300) into at least the first and second hash portions (301, 302) and a third hash portion (303), and the storing unit (53) is arranged to discard each of the third hash portions (303); and
- 5 the database access unit (42) is arranged to divide the search hash value (300) derived from the demanded resource identifier (200) into corresponding at least the first and second hash portions (301, 302) and a third hash portion (303).
3. The resource access filtering system of claim 1, wherein the storing unit (53) is arranged to provide the hash array (410) as an index comprising a set of hash index entries
- 10 (411) based on the first hash portions (301), wherein each of the hash index entries (411) comprises the pointer (412) identifying a respective one of the main data blocks (421) in the main data section (420).
4. The resource access filtering system of claim 3, wherein each of the hash index entries (411) further comprises a length value (413) of the respective main data block (421) in
- 15 the main data section (420), whereby the pointer (412) and the length value (413) together select the respective main data block (421) from the main data section (420).
5. The resource access filtering system of claim 1, wherein the main data section (420) comprises a plurality of records (422), wherein each record (422) comprises a respective one of the second hash portions (302) and the associated category code (310) for a given one
- 20 of the plurality of resource identifiers (200).
6. The resource access filtering system of claim 5, wherein:
- the storing unit (53) is further arranged to provide a page data section (440), and wherein each record (422) in the main data section (420) further comprises a page index (426) which indexes a page data block (444) in the page data section (440) comprising one or more
- 25 page data strings (441) and one or more associated page category codes (311);
- the communication unit (41) is arranged to obtain the demanded resource identifier (200) including a demanded page (204); and
- the database access unit (42) is arranged to receive the demanded page (204) and to compare the demanded page (204) against the page data strings (441) in the respective page
- 30 data block (444) and return the associated page category code (311).

7. The resource access filtering system of claim 6, wherein at least one entry in the page index (426) comprises a default value which indicates that there are no page data strings (441) in the page data section (440) for a particular one of the records (422).
8. The resource access filtering system of claim 1, wherein:
- 5 the storing unit (53) is further arranged to provide a conflict data section (430) comprising the category code (310) and an undivided hash value (300) of at least one of a first resource identifier (200) and a second resource identifier (200) for which the first and second hash portions (301,302) thereof conflict, and wherein at least one of the records (422) in the main data section (420) for the first and second resource identifiers (200) indicates that the
- 10 conflict data section (430) contains the undivided hash value (300); and
- the database access unit (42) consults the conflict data section (430) when the respective record (422) in the main data section (420) for the demanded resource identifier (200) so indicates.
9. The resource access filtering system of claim 1, wherein:
- 15 the filter unit (40) comprises at least a faster-access primary storage device (44) and a slower-access secondary storage device (45); and
- in a first memory-based mode, the entire database structure (400) is loaded into the faster-access primary storage device (44) and is initialised by converting each of the pointers (412) in the main hash array (410) into memory location references which directly or indirectly
- 20 reference memory address locations where the main data blocks (421) of the main data section (420) are loaded, and is accessed in the faster-access primary storage device (44); and
- in a second shared-access mode, the main hash array (410) is loaded into the faster-access primary storage device (44), whilst at least the main data section (420) is retained on
- 25 the secondary storage device (45) and is initialised by converting each of the pointers (412) in the main hash array (410) into offsets from a start of a file of the database structure (400) as stored on the secondary storage (45), and is accessed by obtaining the converted offset (412) from the main hash array (410) in the primary storage device (44) and loading only the required data block (421) from the main data section (420) into the primary storage device
- 30 (44).
10. The resource access filtering system of claim 9, wherein the filter unit (40) is selectively operated in either the first memory-based mode or in the second shared-access mode.

11. The resource access filtering system of claim 1, wherein:
- the generator unit (50) further comprises an updater unit (54) arranged to produce an update structure (500); and
- the filter unit (40) is arranged to apply the update structure (500) to a current copy of the database structure (400a) to produce an updated new database structure (400b) incorporating a plurality of desired updates;
- wherein the update structure (500) comprises at least a main update section (520) comprising a list of update commands (521) each followed by one or more parameters (522) to insert, modify or delete a respective record (422) in the main data section (420) and the update commands (521) are sorted such that the order of the commands (521) corresponds to the order of the records (422) in the main data section (420).
12. The resource access filtering system of claim 11, wherein the update structure (500) includes at least:
- an INSERT command which instructs the filter unit (40) to insert a new record (422) in the main data section (420) at a location identified by a position parameter (522) giving an offset position where the instructed command is to be performed, and also provides parameters (522) for the second hash portion (302) and the category code (310) of the new record (422) and the first hash portion (301) such that the main hash array (410) is recalculated by the filter unit (40) to allow for the new record (422);
- a MODIFY command which instructs the filter unit (40) to modify an existing record (422) in the main data section (420) wherein the parameters (522) provide new data to be written into the database structure (400) to replace the current data of the relevant record (422); and
- a DELETE command which instructs the filter unit (40) to delete a record (422) from the main data section (420) of database structure (400) at a location identified by a position parameter (522) giving an offset position where the instructed command is to be performed, and a parameter (522) comprising the respective first hash portion (301), such that the filter unit (40) recalculates the main hash array (410) to account for the now deleted record (422).
13. The resource access filtering system of claim 12, wherein the position parameter (522) gives a relative position which is relative to a position of a previous command or, if no previous commands have been given, is relative to a default starting position.
14. The resource access filtering system of claim 11, wherein the storing unit (53) is further arranged to provide a page data section (440), and the updater unit (54) is arranged to provide the update structure (500) including a page update section (540) having a list of page

update commands (541) each followed by one or more page update parameters (542) to insert, modify or delete a respective page string (441), and the page update commands (541) are sorted by the first hash portion (301) and then by the page string (441).

15. The resource access filtering system of claim 11, wherein the storing unit (53) is further arranged to provide a conflict data section (430), and the updater unit (54) is arranged to provide the update structure (500) including a conflict update section (530) having a list of conflict update commands (531) each followed by one or more conflict update parameters (532) to insert, modify or delete a respective undivided hash value (300), and the conflict update commands (531) are sorted by the undivided hash values (300).

10 16. The resource access filtering system of claim 11, wherein the filter unit (40) is arranged to update the database structure (400) by copying a current live copy of the database structure (400a) in a plurality of chunks whilst applying the update commands (521) separately in turn to each of the copied chunks to produce the updated new database structure (400b).

15 17. The resource access filtering system of claim 1, wherein the resource identifiers (200) identify resources in a computing network.

18. The resource access filtering system of claim 1, wherein the resource identifiers (200) comprise Uniform Resource Locators.

19. The resource access filtering system of claim 1, wherein the resource identifiers (200) comprise at least a host portion (202) and the hash unit (52) hashes the host portion (202) to provide the hash value (300).
20

20. A generator apparatus (50) to create a database structure (400), comprising:

a categorisation unit (51) arranged to provide a plurality of resource identifiers (200) each associated with a respective category code (310);

25 a hashing unit (52) arranged to hash each of the plurality of resource identifiers (200) to provide a hash value (300), and to divide the hash value (300) into at least first and second hash portions (301, 302); and

30 a storing unit (53) arranged to output the database structure (400) including at least a hash array (410) and a main data section (420), wherein the main data section (420) comprises the second hash portions (302) and the respective category codes (310) divided amongst a plurality of main data blocks (421), and the hash array (410) comprises the first hash portions (301) each associated with a pointer (412) to a respective one of the main data blocks (421) in the main data section (420).

21. A filter apparatus (40) to control access to resources (32) in a computing system, comprising:

5 a filter unit (40) arranged to control access by a client device (10) to a resource (32) with reference to a category code (310) obtained from a database structure (400) according to a demanded resource identifier (200); and

10 a database access unit (42) arranged to obtain the category code (310) appropriate to the demanded resource identifier (200) from the database structure (400), by generating a search hash value (300) based on the demanded resource identifier (200) and dividing the search hash value (300) into at least first and second hash portions (301, 302), identifying a respective main data block (421) in a main data section (420) of the database structure (400) by consulting the hash array (410) using the first hash portion (301), and retrieving the category code (310) from the respective main data block (421) according to the second hash portion (302).

22. A computer readable storage medium having recorded thereon a database structure (400) storing data representing a plurality of resource identifiers (200) each associated with a respective category code (310), comprising:

20 a main data section (420) having a plurality of main data blocks (421), wherein each of the plurality of main data blocks (421) comprises a plurality of second hash portions (302) and each of the plurality of second hash portions (302) is associated with a respective category code (310); and

a main hash array (410) comprising a plurality of first hash portions (301), wherein each of the first hash portions (301) is associated with a reference pointer (412) to a respective one of the plurality of main data blocks (421) in the main data section (420);

25 wherein each of the first hash portions (301) and the second hash portions (302) is divided from a hash value (300) derived a respective one of the plurality of resource identifiers (200).

23. A method of creating a database structure (400), comprising the computer-implemented steps of:

30 providing a plurality of resource identifiers (200) each associated with a respective category code (310);

hashing each of the plurality of resource identifiers (200) to provide a respective hash value (300) and dividing the respective hash value (300) into at least a first hash portion (301) and a second hash portion (302); and

storing at least a main hash array (410) and a main data section (420) in the database structure (400), wherein the main data section (420) comprises a plurality of main data blocks (421) each comprising a set of the second hash portions (302) associated with the respective category codes (310), and the main hash array (410) comprises the first hash portions (301) each associated with a reference (412) to a respective one of the plurality of main data blocks (421) in the main data section (420).

24. A method of retrieving data from a database structure (400), comprising the computer-implemented steps of:

receiving a demanded resource identifier (200) which identifies a resource (32) in a computing system;

creating a hash value (300) based on the resource identifier (200) and dividing the hash value (300) into at least first and second hash portions (301, 302);

identifying a main data block (421) in a main data section (420) of the database structure (400) by consulting a hash array (410) using the first hash portion (301);

and retrieving a category code (310) from the identified main data block (421) according to the second hash portion (302).

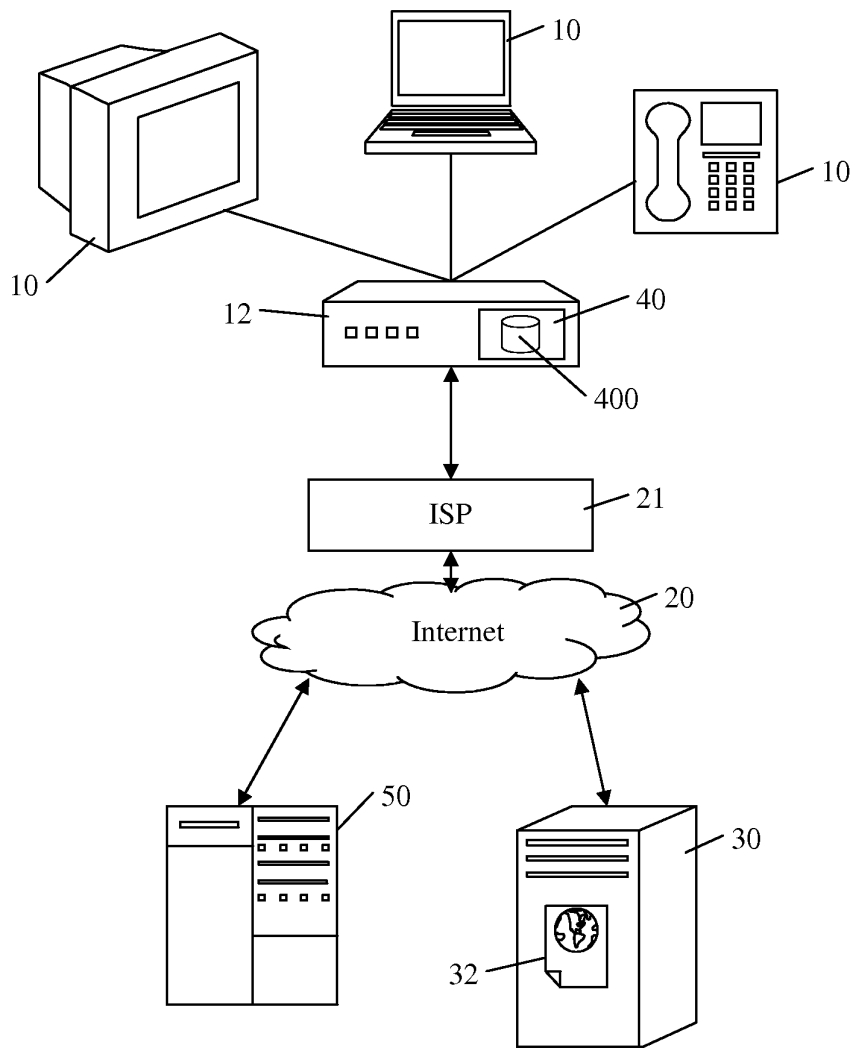


Fig. 1

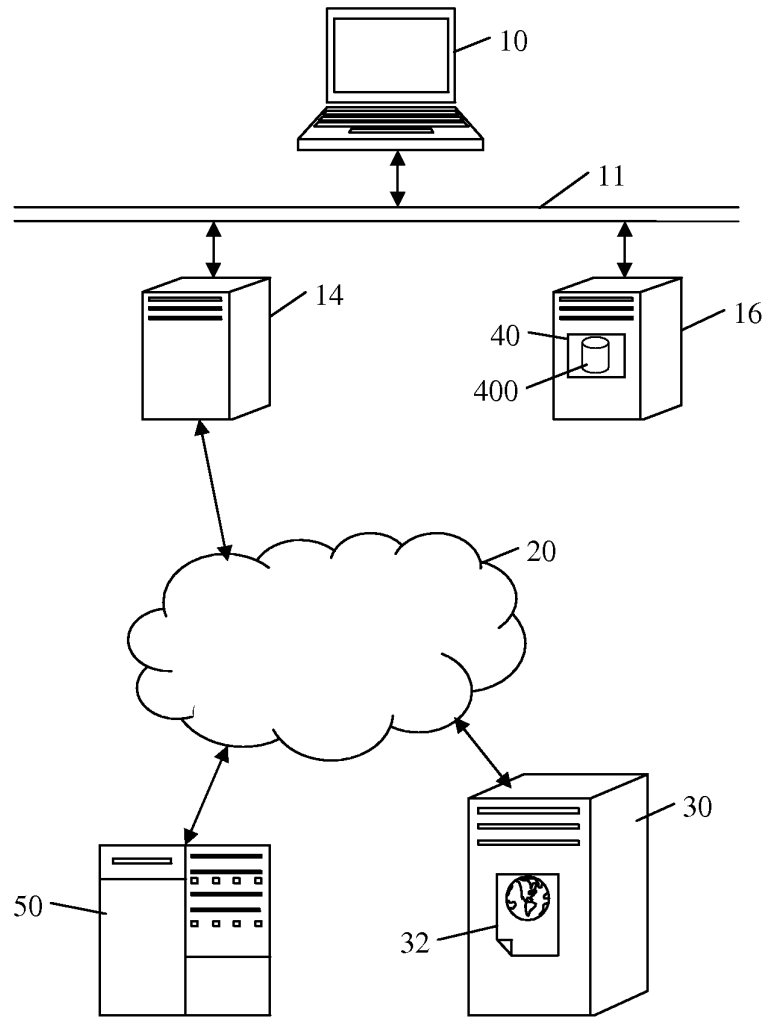


Fig. 2

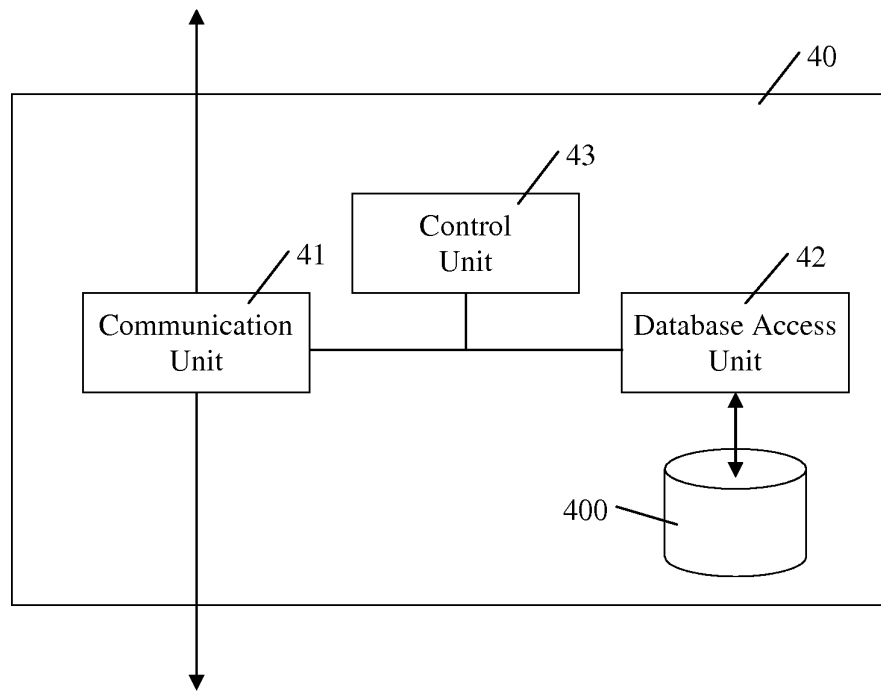


Fig. 3

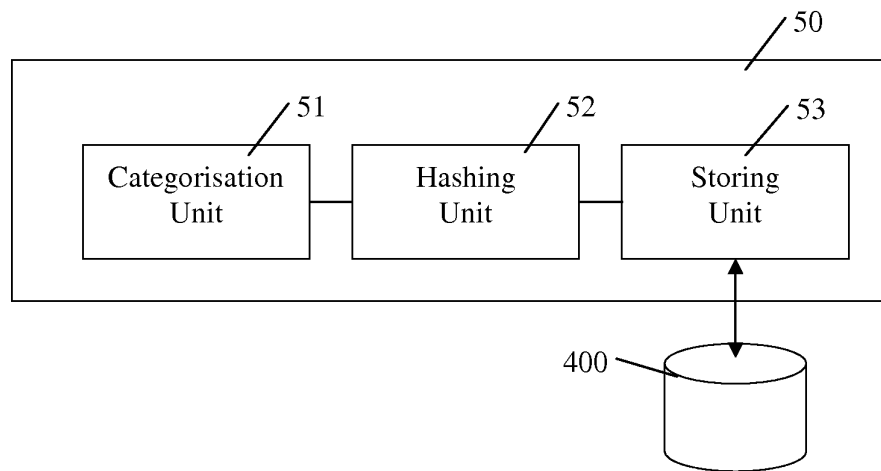


Fig. 4

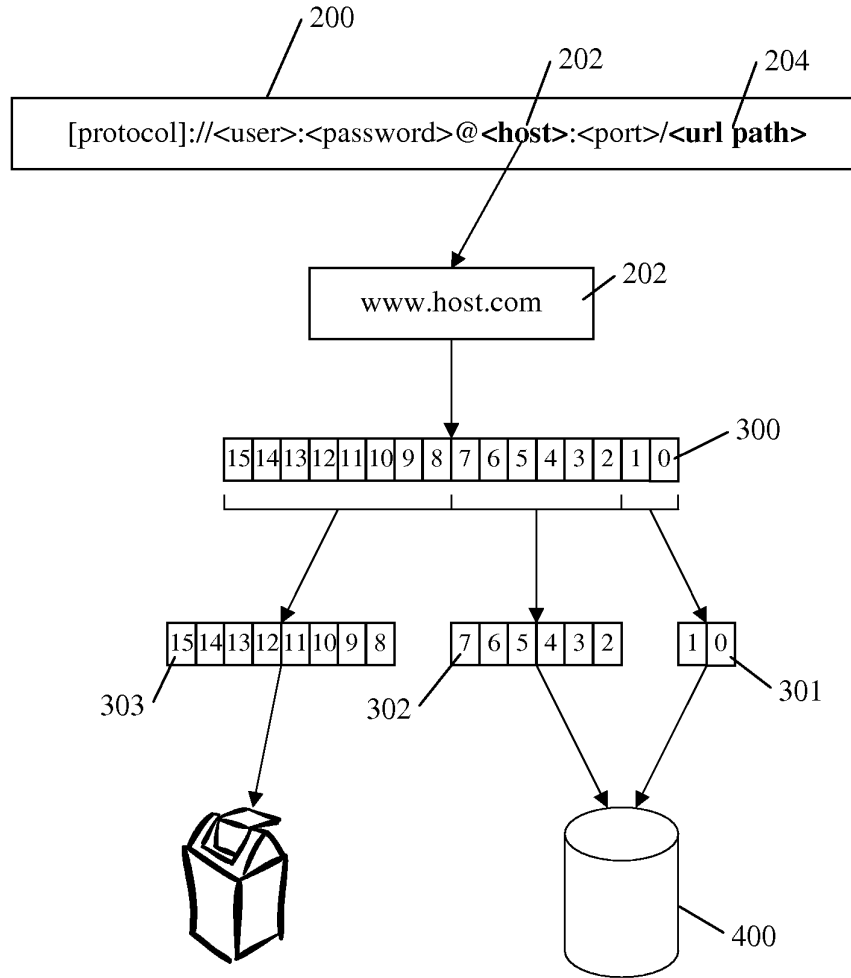


Fig. 5

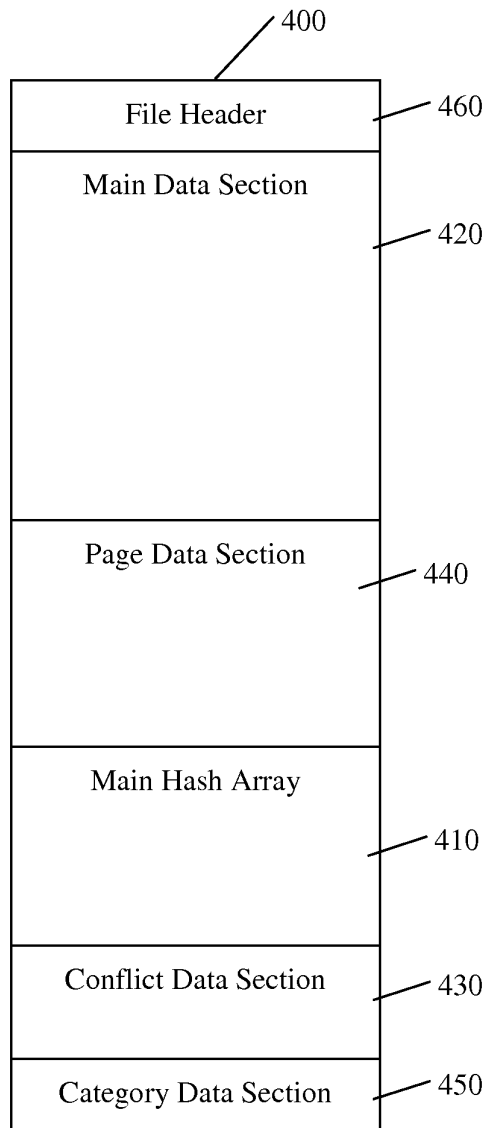


Fig. 6

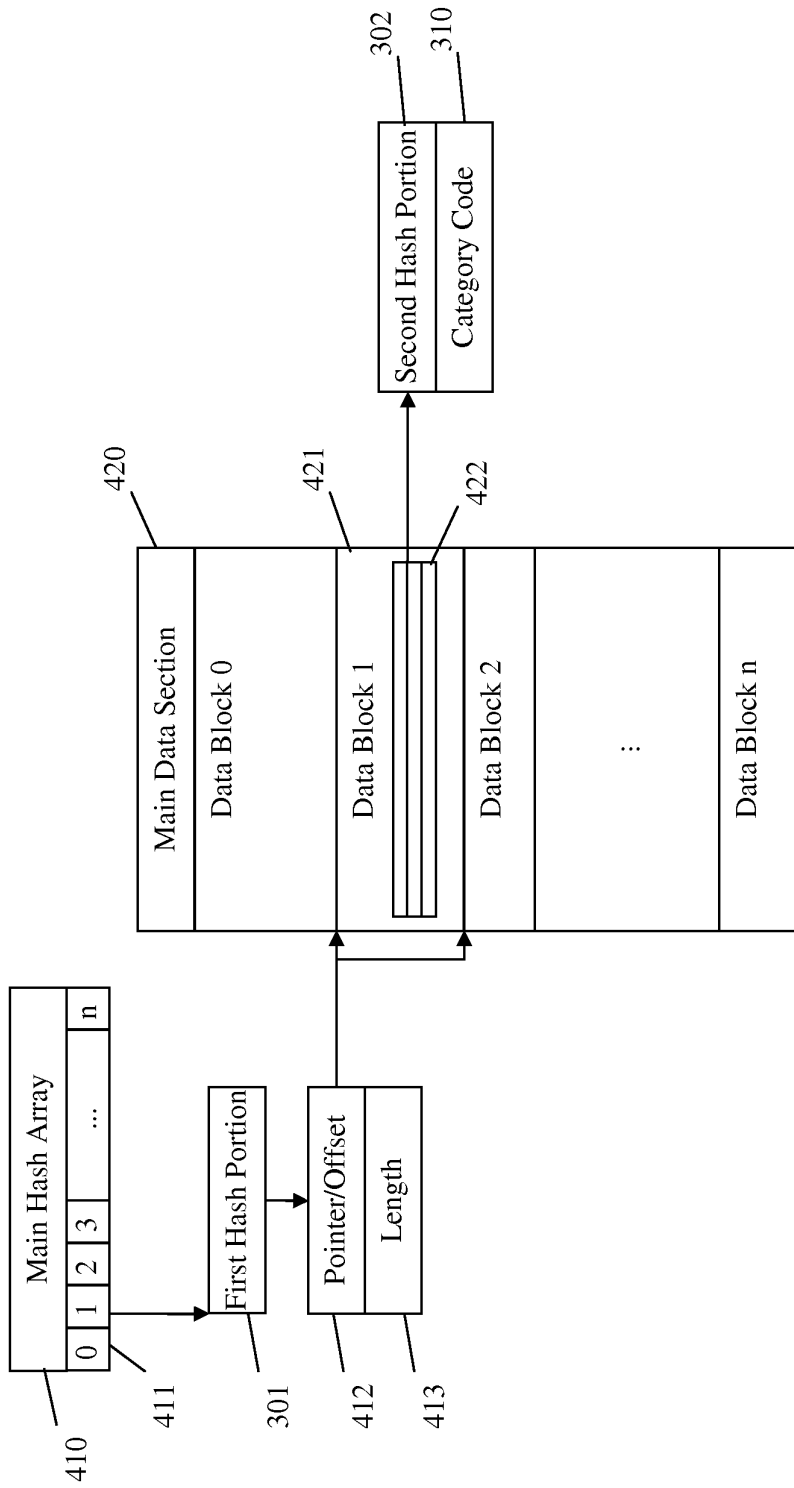


Fig. 7

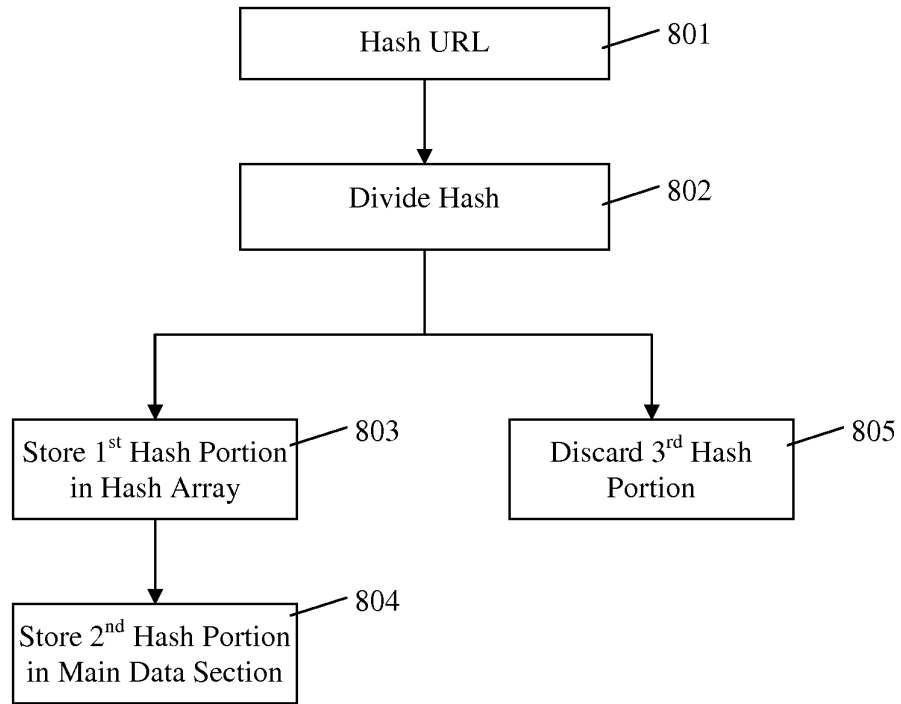
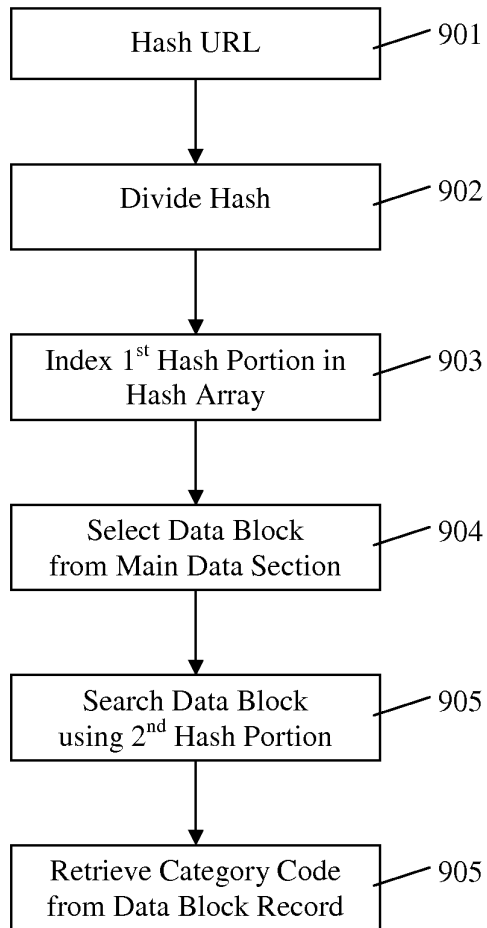


Fig. 8

9/15

**Fig. 9**

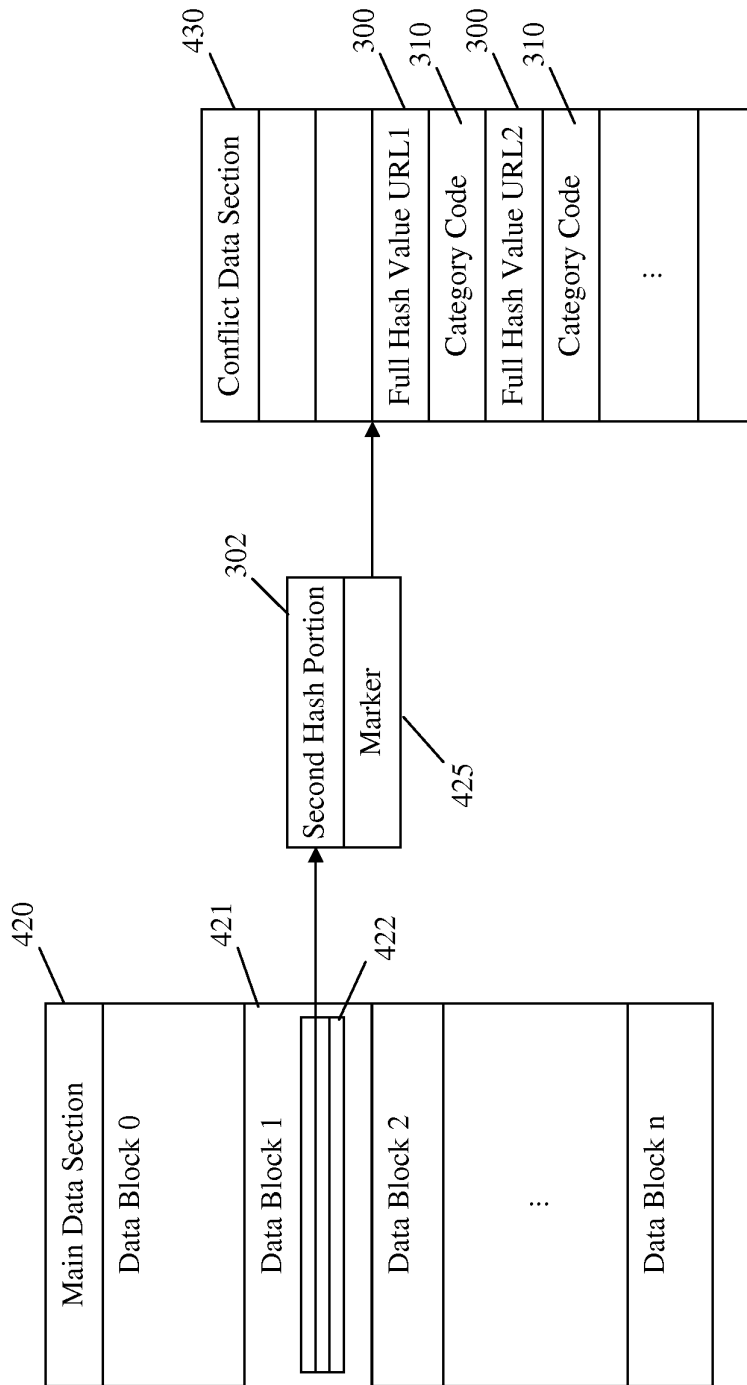


Fig. 10

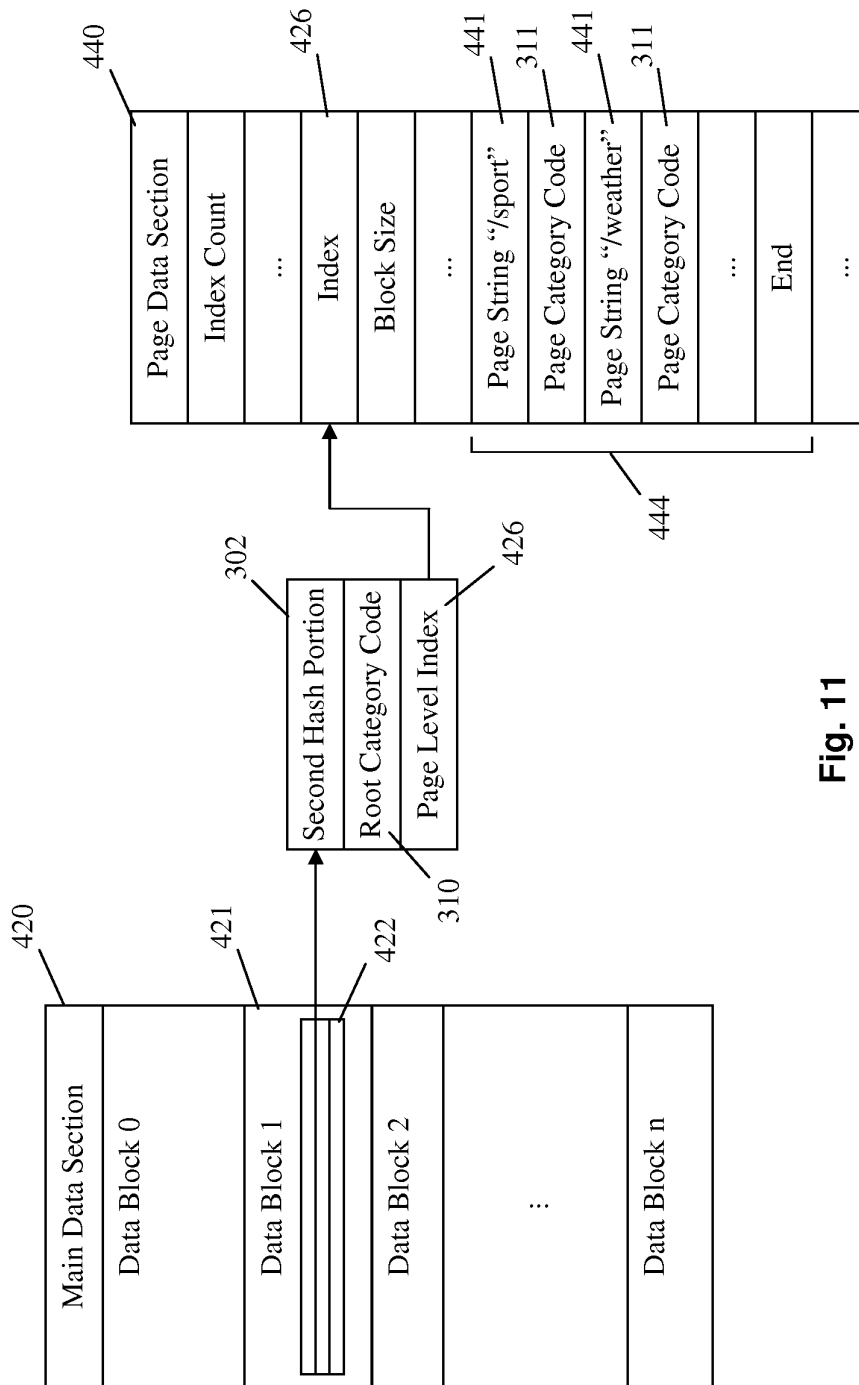


Fig. 11

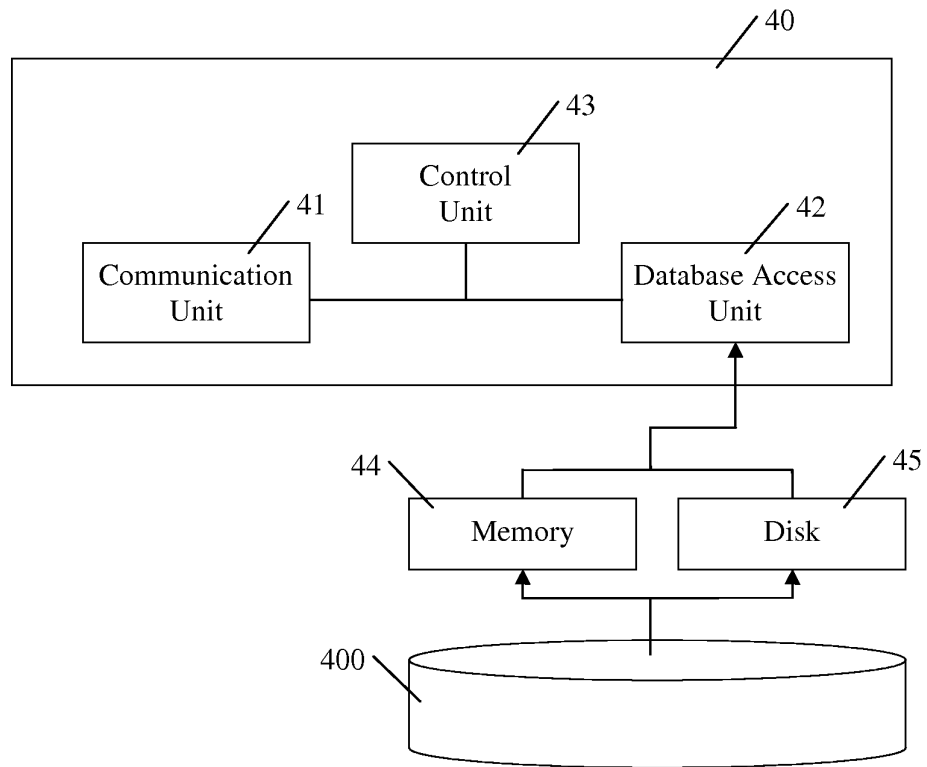


Fig. 12

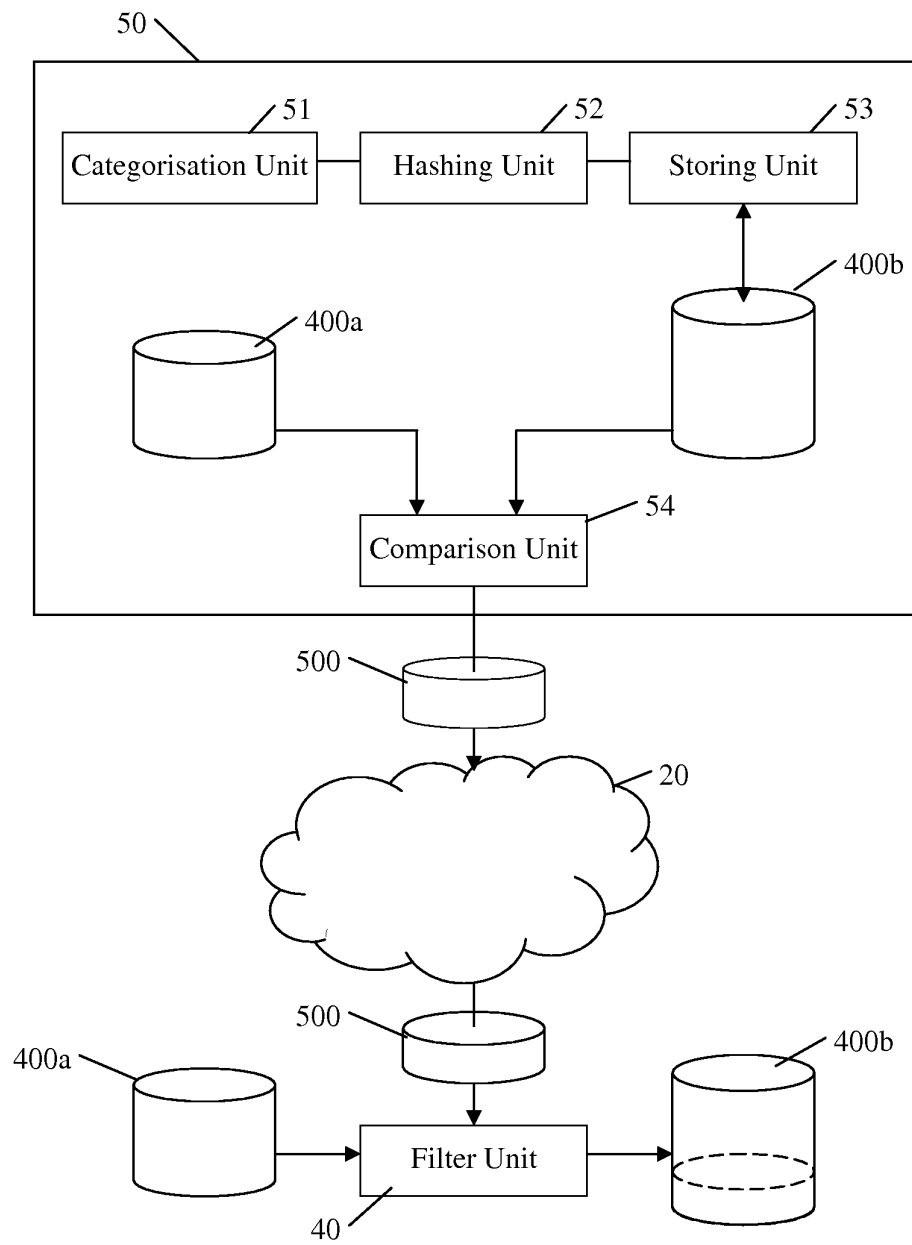


Fig. 13

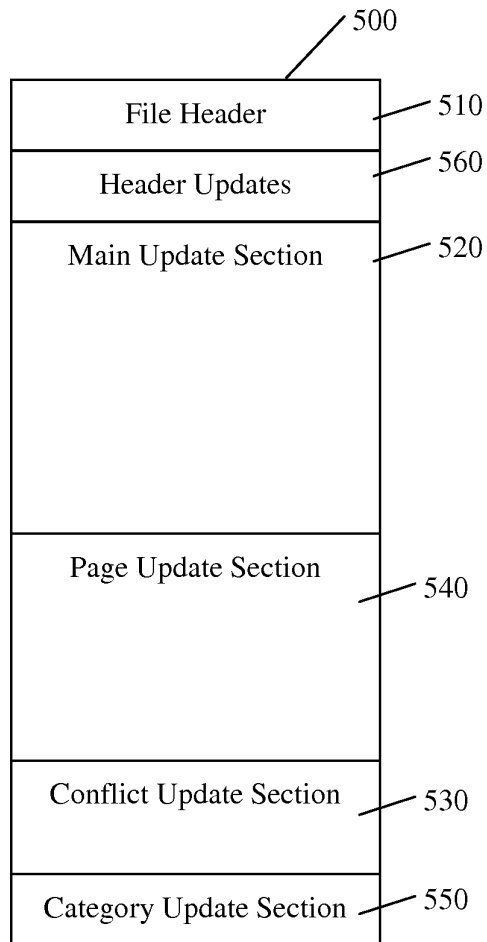


Fig. 14

Main Update Section					
INSERT	1 st Hash Portion	2 nd Hash Portion	CC	PLI	Position
MODIFY			CC	PLI	Position
DELETE	1 st Hash Portion				Position

520

521 522

Fig. 15

Page Update Section				
INSERT	0 / PLI	LGTH	PGE + PCC	Position
MODIFY	0 / SIZE	PCC / -		Position
DELETE	0 / PLI			Position

540

541 542

Fig. 16

Conflict Update Section				
INSERT	Full Hash Value	Category Code	Page Level Index	Position
MODIFY	Category Code		Page Level Index	Position
DELETE	1 st Hash Portion			Position

530

531 532

Fig. 17