27 JAN 2005

METHOD AND APPARATUS FOR ACCESSING WEB PAGES

FIELD OF THE INVENTION

This invention relates to a method and apparatus for accessing web pages, especially web pages delivered over the Internet. In particular, the invention relates to a method and apparatus for permitting access only to suitable web sites.

BACKGROUND ART

The World Wide Web has been growing explosively in recent years and, as is well known, provides a global network of information arranged in the form of web-pages. It is possible to access pages on the world wide web by simply typing in the address of the page, but it is normally much more convenient to use "linking" to navigate from one web-page to another.

A link is an indication on one web page of another web-page. The user can use a graphical pointing device such as a cursor controlled by a mouse to select the link. Then, the web-page referred to by the link is displayed.

To link web-pages, an area of a web-page is determined in the script defining the web-page generally in "HTML" or hypertext mark up language. When a user double clicks using a mouse in the defined area the new web-page selected is displayed. The image to be displayed on the defined screen area to allow the user to identify the link and the address of the new web-page are all stored in the script.

Web-pages are arranged on web-sites, which can be considered to be a group of web-pages grouped together by a single user, company or for any other reason whatsoever. The links on web-pages can relate to web-pages on the same or on other sites.

The global nature of the World Wide Web and the ease with which individuals can provide their own web-pages has resulted in an explosive growth of

content of all types on the World Wide Web. There are pages relating to every subject imaginable including subjects often considered taboo or illegal such as discussions of extreme sexual practices, sites depicting violent or revolutionary material and other sites of a questionable or undesirable nature.

Because of the extreme nature of some of the material there has been an increase in demand for filtering software to restrict access to unsuitable web-sites. Such software for restricting access may be required, for example, by parents to allow their children to safely "surf" the web without coming across unsuitable material. However, there are many other applications where such filtering may be useful. Corporations may wish to restrict the access of their employees. Other users may simply find some Internet content offensive and unacceptable and will not wish to have access to such content.

In order to allow access only to suitable web-sites some means of labelling or classifying web-sites needs to be found. The vast number of web-sites makes this a very difficult task. In the past, a number of approaches have been taken.

In one approach, a web browser allows access only to those web-sites listed on a list of approved web-sites prepared by a central ratings service. The ratings service provides the list which includes a large number of web-sites, by one or many criteria. For example, web-sites may be rated for the levels of violence and sexual content. A web browser may then permit access only to those web-sites that are suitably rated.

The problem is that in view of the vast number of sites and the many changes being made on a daily basis it is inevitably the case that many sites will be un-rated.

Access to un-rated web-sites can be simply completely blocked. Another approach is to attempt to deal with un-rated web-sites using some form of automatic filter. Such filters may work, for example, on word-recognition software. Unfortunately, such filters are very unreliable.

Instead of a central service rating web-sites an alternative is to use self-rating approaches in which web-sites voluntarily rate themselves for content. The disadvantage here is that many users will be unwilling or will simply not bother to rate their own web-site. Moreover, in view of the global nature of the Internet it is inevitable that users in different countries or regions will have very different ideas and standards as to what is acceptable. Thus, it is unlikely that self-rating can ever be made to work properly.

An alternative approach is to identify and create a black list of "offensive" or "unfavourable" sites. A proxy server may then be used to block access to such web-sites and allow access to all others. However, this requires the tedious process of identifying all unsuitable web-sites. The disadvantage lies in the inability to identify new web-sites which should be banned, giving the explosive growth in the number of web-sites.

A further reason to wish to restrict web-sites is for the purpose of providing access only to web-sites on a particular topic, rather than avoiding web-sites with unsuitable content.

Accordingly, there is a real and urgent need for an improved approach to providing web access only to suitable web-sites that addresses the difficulty with un-rated sites.

SUMMARY OF THE INVENTION

According to a first aspect of the invention, there is provided a method of accessing web-sites made up of at least one web page, including

providing a suitability list of web-sites where a level of suitability is associated with each web-site in the suitability list, the level of suitability indicating the number of links that may be followed from the web-site,

allowing access to an initial web-site selected from those present on said suitability list,

allowing access to secondary web-sites no more than an allowed number of links from a web-site on said suitability list, the allowed number of links being determined by the suitability level of said web-site on said suitability list, and

barring access to other web-sites.

The method may initially only allow access to the web-sites on the list and allow access to secondary web-sites only by accessing links. Alternatively, direct access to secondary web-sites may be permitted if it can be determined that the secondary web-site is indeed within the allowed number of links from a web-site included on the suitability list.

The suitability list may be a list of trusted sites with acceptable content or a list of sites relevant to a particular topic.

The suitability list contains a list of web-sites and a level of suitability associated with each web-site. The level of suitability may be conveniently stored in the form of an integer. For example, the integer "9" may be initially assigned to those web-sites rated as safe. Alternatively, different integers may be assigned to different web-sites. A further alternative is to simply list the web-sites on the suitability list and assign the same level of suitability to each.

The integer associated with an accessed web-site may be reduced every time a link is followed. The integer may be decremented by 1, or an amount determined from some algorithm, e.g. geometric decrement. This may be implemented by storing in a variable the integer indicating the level of suitability of the presently accessed web-site. Each time a link is followed, this integer may be decreased. When the integer reaches a predetermined value, conveniently 0, no further linking may be allowed.

It is possible to speed access to secondary sites not initially on the suitability list by automatically adding new sites to the suitability list. New sites at a given number of links away from a site on the suitability list may be added to the suitability list at a level of suitability calculated as the level of suitability of the site on the

suitability list minus the number of links needed to reach the new site. Thus, if a web-site is located in the suitability list as having a level of suitability of 7, all those web-sites reachable by two links from that web-site may be assigned a level of suitability of 5.

This may in particular be useful when the suitability list is to relate to sites having a particular content. In this way, the suitability list may be built up to be a limited index based on a topical interest, such as music, the arts, science or technology.

A data base of sites may built into a proxy server controlling Internet access to limit access to the Internet.

A web-crawler may be provided to search using a non-linear search for candidate web-sites linked to from web-sites already in the suitability list. If the candidate web-site is checked and found to be suitable it may be added to the suitability list as mentioned above.

The invention also envisages a method of permitting a user to access web-sites each comprising at least one web page, wherein a suitability list of web-sites is provided, the method including

accepting an input indicating a web page,

determining if access to the web-site including said input web page is allowed, wherein access is allowed to an initial web-site selected from those present on said suitability list, or to secondary web-sites no more than an allowed number of links from a web-site on said suitability list, the allowed number of links being determined by the suitability level of said web-site on said suitability list, and access is barred to other web-sites, and

displaying the input web page if it is from a web site to which access is allowed.

The method may restrict access to internet services on sites with a level of

suitability below a predetermined restriction level.

The invention also includes a computer program for allowing a user to access web-sites using a computer system having a display, a memory store for storing a suitability list, an input device and an internet connection, wherein a level of suitability is associated with each web-site in the suitability list, the level of suitability indicating the number of links that may be followed from the web-site, the computer program being operable to cause the computer system to carry out a method including the steps of

accepting input from the user indicating a web page to be displayed,

allowing access through the internet connection to an initial web-site selected from those present on said suitability list,

allowing access through the internet connection to secondary web-sites no more than an allowed number of links from a web-site on said suitability list, the allowed number of links being determined by the suitability level of said web-site on said suitability list,

displaying on the display the web-page input by the user if it is on a web-site to which access is allowed, and

barring access to other web-sites.

The invention also envisages a computer system including

a display,

a memory store for storing a suitability list of web-sites wherein a level of suitability is associated with each web-site in the suitability list, the level of suitability indicating the number of links that may be followed from the web-site,

an input device for accepting user input,

an internet connection for connecting to the internet,

code to accept from the input device input indicating a web page to be displayed,

code to allow access through the internet connection to a web-site selected from those present on said suitability list or from other web-sites no more than an allowed number of links from a web-site on said suitability list, the allowed number of links being determined by the suitability level of said web-site on said suitability list,

code to display on the display the web-page input by the user if it is on a web-site to which access is allowed, and

code to bar access to other web-sites.

The system may be a single computer or a networked computer system. In particular, it may often be convenient to store the suitability list remotely from the user, for example on a proxy server.

The invention may identify a wider area of web-sites that have not been censored than is possible by simply a human rating web-sites alone. The underlying concept lies in the analogy of human trust. We trust the friends of our friends, but not as much as our friends.

An important advantage is the ability to self-identify the trustworthiness of unknown web-sites. Every Internet service provider and private proxy server may build their data base of circle of suitability. Proxy servers can cooperate to exchange or share their data base or data bases by increasing infinitely the coverage of web-sites.

BRIEF DESCRIPTION OF THE DRAWINGS

For a better understanding of the invention, and purely by way of example, specific embodiments will now be described with reference to the accompanying drawings in which

Fig.1 illustrates the concept of decreasing level of suitability,

Fig.2 illustrates a computer system according to the invention,

Fig.3 shows an illustration of the arrangement of information and the information flow,

Fig. 4 illustrates a specific embodiment of the invention, and

Fig. 5 illustrates a modification of the embodiment of Figure 4 in which some services are denied at low levels of suitability.

DETAILED DESCRIPTION

Referring to Fig.1, the basic concept of the invention is illustrated. The drawing shows a staircase with the level of suitability on the vertical axis. Initially, a web-site is accessed on the suitability list and has a high level of suitability. There is free accesses to sites and links at this level of suitability. Every time a link is followed, the level of suitability decreases by 1, indicated on the drawing by moving one step to the right. After a predetermined number of steps, the suitability level reduces to 0 at which point there is no access to links.

Fig.2 shows a computer system 11 according to the invention. The computer system 11 has a display 13, a keyboard 15, a mouse 17, a memory store 19, a processor 20 and a network connection 21. The keyboard 15 and mouse 17 function as input devices.

The network connection 21 may be connected to a proxy server 23 that allows access to the Internet; accordingly the network connection 21 has the functionality of a Internet connection. A firewall 25 may be provided having variables 27 defining the level of permitted access to the Internet.

The invention may be implemented by using a proxy server to allow access to the Internet as is known. The proxy server may include the suitability list database, which contain the list of known web-site addresses and the associated

level of suitability.

Fig.3 is a conceptual illustration of browsing the Internet. Starting at a known web-site 1, the user may browse the web-site at 3. When a new web-site is encountered, the database 5 of known web-sites is checked and a value returned relating to whether the web-site address is known and the level of suitability of that web-site. If appropriate, i.e. if the new web-site is on the suitability list or not more than a predetermined number of links away from a web-site on the suitability list, the known web-site at 1 is updated to become the new web-site and Internet browsing at 3 permitted as before.

A description of a method according to an embodiment of the invention will now be presented with reference to Fig. 4.

Firstly, the suitability level L is initialised to zero (step 501). A web page is selected by a user (step 503) by entering a universal resource locator (URL), and it is determined (step 505) whether the URL has been entered directly. If so, L is set to zero (step 507), otherwise, L is decremented by one (step 509).

The URL address is then extracted (step 511) by stripping sub-directories, parameters and the like, to determine whether the URL is on the suitability list (step 513).

If not, a test determines whether L is greater than zero (step 515); if not a message is displayed in step 517 reading "Unsuitable site, access denied". If L is greater than zero, then the web address and L are added to the suitability list (step 519) and the web page displayed (step 521).

If the URL is on the suitability list, the suitability level on the suitability list, L_{List} is retrieved (step 523), and if L is greater than L_{list} (step 525) the suitability list is updated by replacing L_{list} by L (step 527). Otherwise, the value of L is updated to be L_{list} (step 529). Either way the web page is then displayed (step 521).

After the web page is displayed processing may return to step 503 to select a new web page.

27 JAN 2005

In a further preferred embodiment of the invention reduced services are available at low, but non-zero levels of suitability. In the embodiment, a value L_{min} is predetermined that sets a suitability level below which some web services are to be unavailable.

Figure 5 illustrates a process that replaces step 521 in the method of Figure 4. Instead of simply displaying the web page further processing is performed.

It is tested whether L is less than a predetermined restriction level L_{min} in step 601. If not, full access to the web site is granted (step 603).

Otherwise, the firewall 25 is updated to restrict web access. This is done by updating data 27 stored in the firewall to deny access to predetermined services. These may include Internet Relay Chat services (IRC), File Transfer or Download services (FTP), https services, i.e. using a secure socket layer (SSL) typically for a transaction, or special features by disabling specific ports.

It is also possible to bar access to sub-directories of risky web-sites. In step 607, access to sub directories is tested for and a web page displayed (step 603) only to the main page.

If access is to be denied, a message may be displayed (step 609) to indicate "High risk site, services denied".

In order to reset the firewall connections to allow full access to suitable sites, a step 611 may be added between steps 523 and 525 of Figure 4 to check whether the new site from the suitability list is sufficiently suitable and if so to reset the firewall to allow full access.

The above specific embodiments are described purely by way of example and the skilled person will realise that modifications may be made. In particular, any suitable computer hardware or other apparatus may be used to deliver internet access. The flow charts may be implemented by program code, stored permanently or temporarily in the computer or downloaded as required by a network.

In the embodiments described above the suitability list is stored in a memory of a computer. However, the suitability list may also be downloaded as required, or even calculated in real time.

Furthermore, the invention has application in the situation where it is desired to restrict web-sites to web-sites relevant only to a particular topic, to a particular level of trust, or any other application where access only to a subset of sites is required.

CLAIMS

1. A method of accessing web-sites, including

providing a suitability list of web-sites where a level of suitability is associated with each web-site in the suitability list, the level of suitability indicating the number of links that may be followed from the web-site,

allowing access to an initial web-site selected from those present on said suitability list,

allowing access to secondary web-sites no more than an allowed number of links from a web-site on said suitability list, the allowed number of links being determined by the suitability level of said web-site on said suitability list, and

barring access to other web-sites.

2. A method according to claim 1 wherein

access is initially allowed only to web-sites on said suitability list and

secondary web-sites may only be accessed by following a link.

3. A method according to claim 1 wherein the suitability list contains a list of web-sites and a level of suitability associated with each web-site.

4. A method according to claim 3 wherein the level of suitability is stored as an integer.

5. A method according to claim 1 including

storing the suitability level of an accessed web-site in a level variable,

decrementing the level variable every time a link is followed to a web-site not on said suitability list, and

preventing access to linked web-sites when said level variable is less than or equal to a predetermined value.

6. A method according to claim 1 further including

automatically adding sites to the suitability list that are linked to by a site on the suitability list,

the new sites at a given number of links away from a site on the suitability list being added to the suitability list with a level of suitability equal to than that of said site on the suitability minus said given number.

7. A method according to claim 6 including using a web-crawler to search for candidate web-sites linked from web-sites already in the suitability list, and adding the candidate web-sites to the suitability list if the candidate web-sites are linked to by web-sites on the suitability list using no more than a number of links determined by the suitability level of said web-site on the suitability list.

8. A method of permitting a user to access web-sites each comprising at least one web page, wherein a suitability list of web-sites is provided, a level of suitability being associated with each web-site in the suitability list, the level of suitability indicating the number of links that may be followed from the web-site, the method including

accepting an input indicating a web page,

determining if access to the web-site including said input web page is allowed, wherein access is allowed to initial web-sites selected from those present on said suitability list, or to secondary web-sites no more than an allowed number of links from a web-site on said suitability list, the allowed number of links being determined by the suitability level of said web-site on said suitability list, and access is barred to other web-sites, and

displaying the input web page if it is from a web site to which access is allowed.

9. A method according to claim 8 wherein

access is initially allowed only to web sites on said suitability list and

access is subsequently allowed to web-sites not present on said suitability list only by following a link.

10. A method according to claim 8 wherein the suitability list contains a list of web-sites and a level of suitability associated with each web-site.

11. A method according to claim 10 wherein the level of suitability is stored as an integer.

12. A method according to claim 8 including

storing the suitability level of an accessed web-site in a level variable,

decrementing the level variable every time a link is followed to a web-site not on said suitability list, and

preventing access to linked web-sites when said level variable is less than or equal to a predetermined value.

13. A method according to claim 8 further including

automatically adding sites to the suitability list that are linked to by a site on the suitability list,

the new sites at a given number of links away from a site on the suitability list being added to the suitability list with a level of suitability equal to than that of said site on the suitability minus said given number.

14. A method according to claim 13 including

searching for candidate web-sites linked from web-sites already in the suitability list using a web-crawler, and

adding the candidate web-sites to the suitability list if the candidate web-sites are linked to by web-sites on the suitability list by no more than a number of links determined by the suitability level of said web-site on the suitability list.

15. A method according to claim 8 including the step of restricting access to internet services on sites with a level of suitability below a predetermined restriction level.

16. A computer program for allowing a user to access web-sites using a computer system having a display, a memory store for storing a suitability list, an input device and an internet connection, wherein a level of suitability is associated with each web-site in the suitability list, the level of suitability indicating the number of links that may be followed from the web-site, the computer program being operable to cause the computer system to carry out a method including the steps of

accepting input from the user indicating a web page to be displayed,

allowing access through the internet connection to an initial web-site selected from those present on said suitability list,

allowing access through the internet connection to secondary web-sites no more than an allowed number of links from a web-site on said suitability list, the allowed number of links being determined by the suitability level of said web-site on said suitability list,

displaying on the display the web-page input by the user if it is on a web-site to which access is allowed, and

barring access to other web-sites.

17. A computer program according to claim 16 wherein the computer program causes the computer system

to initially allow access only to web sites on said suitability list and

to allow access to other web-sites only by following a link.

18. A computer program according to claim 16 wherein

the suitability list contains a list of web-sites and a level of suitability associated with each web-site.

19. A computer program according to claim 18 wherein the level of suitability is stored as an integer.

20. A computer program according to claim 16 for controlling the computer to carry out the additional steps of

storing the suitability level of an accessed web-site in a level variable,

decrementing the level variable every time a link is followed to a web-site not on said suitability list, and

preventing access to linked web-sites when said level variable is less than or equal to a predetermined value.

21. A computer program according to claim 20 for controlling the computer to carry out the additional steps of

automatically adding sites to the suitability list that are linked to by a site on the suitability list,

the new sites at a given number of links away from a site on the suitability list being added to the suitability list with a level of suitability equal to than that of said site on the suitability minus said given number.

22. A computer program according to claim 21 for causing the computer system

to use a web-crawler to search for candidate web-sites linked from web-sites already in the suitability list, and to add the candidate web-sites to the suitability list if the candidate web-sites are linked to by web-sites on the suitability list using no more than a number of links determined by the suitability level of said web-site on the suitability list.

23. A computer program according to claim 20 including code to restrict access to internet services on sites with a level of suitability below a predetermined restriction level.

24. A computer system including

a display,

a memory store for storing a suitability list of web-sites wherein a level of suitability is associated with each web-site in the suitability list, the level of suitability indicating the number of links that may be followed from the web-site,

an input device for accepting user input,

an internet connection for connecting to the internet,

code to allow access through the internet connection to an initial web-site selected from those present on said suitability list or from other web-sites no more than an allowed number of links from a web-site on said suitability list, the allowed number of links being determined by the suitability level of said web-site on said suitability list,

code to display on the display the web-page input by the user if it is on a web-site to which access is allowed, and

code to bar access to other web-sites.

25. A computer system according to claim 24 including

code to initially allow access only to web sites on said suitability list and to allow access to other web-sites only by following a link.

26. A computer system according to claim 25 wherein

the suitability list contains a list of web-sites and a level of suitability associated with each web-site.

27. A computer system according to claim 26 wherein the level of suitability is stored as an integer.

28. A computer system according to claim 24 further comprising

code for storing the suitability level of an accessed web-site in a level variable,

code for decrementing the level variable by 1 every time a link is followed to a web-site not on said suitability list, and

code to prevent access to linked web-sites when said level variable is less than or equal to a predetermined value.

29. A computer system according to claim 24 further comprising

code to automatically add sites to the suitability list that are linked to by a site on the suitability list, the new sites at a given number of links away from a site on the suitability list being added to the suitability list with a level of suitability equal to than that of said site on the suitability minus said given number.

30. A computer system according to claim 29 for causing the computer system to use a web-crawler to search for candidate web-sites linked from web-sites already in the suitability list, and to add the candidate web-sites to the suitability list if the candidate web-sites are linked to by web-sites on the suitability list using no more than a number of links determined by the suitability level of said web-site on the suitability list.

27 JAN 2005

ABSTRACT

METHOD AND APPARATUS FOR ACCESSING WEB PAGES

The invention relates to an improved method of browsing web-sites that permits the exclusion of undesirable web-sites. Access is permitted to web-sites on a list of web-sites, and access is also permitted to web-sites no more than a predetermined number of links away from web-sites on the list. The list of web-sites may be added to automatically by adding web-sites linked to by sites already on the list.

(Fig.3)