



## State of Israel

Ministry of Finance - Office of the Accountant-General

Government Internet Committee

WWW.ITPOLICY.GOV.IL

Document: information filtering in the Internet

Written: 17/7/97

updated: 27/7/97

Version: 1.3

Author: Boaz Chen

---

### **Background**

The Internet, since its inception, has been global and “anarchic”, promoting a wide range of activities and freedom of expression and information. On the other hand, as it has expanded and millions of new subscribers have been added, the problematical issues associated with this policy have surfaced. This is expressed by the type of sites, discussion groups and IRC channels that include hard pornography or incitements to violence, racism, terror and others. Attempts to pass laws against the criminal use of the Internet (based on existing laws for other types of media) have met with strong opposition as well as problems of enforcement.

The issue has become the topic of community discussion due to the increasing numbers of children and teenagers who use the Internet and with the intention of protecting them from information of this nature.

### **The Problem**

Attempts to bridge the gap between freedom of expression and conservation of the values we cherish have resulted in a search for solutions that will enable filtering of information.

The requirements for filtering information are different and varied, for example:

- Protecting children and youth from pornography, violence, racism etc.

- Prohibiting entry to amusement sites from the workplace.

- Prohibiting entry to types of services or sites during peak hours.

- And so on...

### **Mechanized Filtering**

The original solutions put forward were based on analysis of information during attempts to enter the site. This refers to programs that check the information before it is displayed to the user, to ensure it meets certain prescribed standards. For example:

programs which scan the text to ensure the non-appearance of specific words or programs which check for the amount of uncovered flesh before displaying pictures.

Solutions such as those have a number of difficulties:

The solutions are specific to particular criteria and will not be compatible with information transferred or presented in a different format. For example, undesirable text that is displayed as a picture, program or even in a different language, will not be effectively filtered.

Mechanical filtering is lacking because of the difficulty in properly defining cross-sections and the overall context. Here are a number of examples:

Programs that prevent the entry to a site containing the word SEX will also prevent academic information on the subject and even sites which contain sexual advice for teenagers. On the other hand, sites that only contain pornographic pictures will not be filtered out.

Attempts to prohibit entry to racially prejudiced sites will often inhibit attempts to enter anti-racially sites as well.

Programs that filter pictures will also filter out artistic images as well as medical, instructional pictures etc.

Checking of information at the time of loading slows down the data acquisition process.

The natural difficulty in defining filtering thresholds and the intense difficulties in translating them into programs, are made harder due to the difference in approaches and cultural backgrounds of the peoples in the virtual world.

### **The Classification and Labeling Solution**

An innovative approach to this problem, which we will discuss in detail in this survey, is based on human classification and evaluation of information.

The intention is, in order to ease the filtering process, to check the information, grading and labeling it based on predefined criteria and a format which enables mechanized use of these labels.

This is similar to the classification methods already in use in the fields of movies and videocassettes, newspapers or books, except that the programs that will base their filtering criteria on labeling will enable monitoring and enforcement while making the filtering suited to the user.

In general terms, there are three elements involved in the process: the distributor of the information, the evaluator responsible for classification based on standards and the user whose filtering program checks the information's labeling and decides whether or not to allow its retrieval. This is shown in the diagram:

This approach, which solves many of the problems associated with the automatic analytical tools, creates new dilemmas and even the suspicion that it is creating censorship. The major questions it raises are:

Who is responsible for the classifications?

According to what categories will the information be classified?

What will be the criteria used for grading within each level?

Who will set the filtering parameters?

These questions and others made the World Wide Web Consortium<sup>1</sup> to set new standard called PICS (Platform for Internet Content Selection)<sup>2</sup>.

The standard defines two things:

Standard label structure - technical structure that defines the format of the label.

A standard structure for the labeling server - definition of how the labeling service provider defines the categories upon which labels are based, the various levels of division within categories and the interface for presentation of grading criteria.

These definitions are completely technical and as such enable disconnection between the world of content of the classification and the technical realization of it. (Similar to what the HTML standard has done for www pages).

This definition, enables different authorities to make classifications without the dependence on filtering programs, and vice versa. Software developers will be able to develop filtering mechanisms without having to comply with a specific classification authority or standard.

In this way, the standard planners hope to maintain the atmosphere of democracy in the Internet, so every user will be able to choose the software, classification authorities, sorting category and levels in each category he wishes to use.

For example:

A user will be able to connect with an authority specializing in the fight against racism and choose which level in each category constitutes a limitation, while making use of the criteria distributed by that authority. Such a choice might look like the following:

Category	Level	Criteria
Type of publisher	3	Institutions and organizations, no private information
History	5	Show all historical fact
Language	2	Enable condemnatory nicknames and jokes, but stop at that.

The information labeled by this authority will be displayed to the user only if it meets the standards (it is possible to define how to deal with totally unlabeled information).

---

<sup>1</sup> The World Wide Web Consortium (W<sup>3</sup>C for short) is a joint venture of industry and universities in the USA, Europe and Asia, for development of protocols.

<sup>2</sup> For details see [www.w3.org/pics](http://www.w3.org/pics) or [www.w3.org/pub/www/pics](http://www.w3.org/pub/www/pics)

## **Labels**

The label is a formatted information, that is to say, information that is readable by a computer and therefore usable by a filtering program. The standard does not define the types of information in the label and, by this, it enables labeling according to different categories. For an example of a technical application of a label, see Appendix A.

Labels may be used in two main ways:

Embedded in the information page - the label may be included as part of the page header. This is applicable to those who wish, to cooperate with the classification authority (a good reason for doing so would be the fact that in most filtering programs it is possible to disable access to all information which has no classification).

Separated - It is also possible to keep groups of labels, which relate to information all across the Internet, in different servers separate from the information. The technical implementation of this comes when, at the time of accessing the information, the filtering program also checks the server containing the labels and searches to confirm the existence of the appropriate label. In this way, classification of information whose distributors do not want to cooperate can also be achieved.

The advantage of the first method is that the classification function can be carried out as a self-initiated or even independent basis and changing the location of the information does not influence it. The authenticity of the label can be confirmed by use of a digitized signature on the label, by its creator.

The second method has two advantages:

Independence between the publisher of the information and its various evaluators. The publisher of the information does not have to do anything for the information to be labeled. (Or even to know that it has been done).

Prevention of increasing in size of all files containing information. This is especially important when there are a large number of evaluators. That is to say, rather than adding a large number of labels to the information, each user can take the labels he requires from the appropriate classifying server. The ability to manage this procedure can be increased by concentrating all these labels together and copying them all into the local server (caching). This has the added advantage of enabling storage of local labels that are sensitive to the language and norms of the user's location.

Every label is related to a URL and therefore labels can refer to information in sites and HTML pages, FTP sites and specific files, CHAT servers, rooms and notices posted on them and the like.

## **Classification**

Classification can be done by the information's distributor (by independently adding a label to the document) or by an external authority.

In addition to defining the label format, the PICS also defines the method of transfer of data between classification servers and the programs that use them. The server contains a vocabulary that defines the types of categories that are used to label sites (every "word" constitutes an evaluation dimension for the site being labeled). For a technical definition of the classification element, see Appendix A.

In this manner, every program will be able to work with every server, by defining the categories of external labeling in it.

For example: The Recreational Software Advisory Council<sup>3</sup> which classifies computer games, decided to classify them according to five levels of each of the following dimensions: violence, nudity, sex and foul language.

The classifiers are generally experts in a particular field (software, games, databases, etc.) or specific types of dangers (racism, violence, fraud, etc.). It is reasonable to expect that authorities that are already doing similar type of classification in different environments will probably provide some of the classification services.

The two types of classifier can be distinguished by:

Professionals - companies or organizations who employ subject specialists e.g. SafeSurf<sup>4</sup>.

Democrats - authorities that classify based on evaluations from other users. This method enables the receipt of evaluation of information based on what similar thinking people have said about it (based on similarities with parallel evaluations).

## **Filtering**

The filtering programs are those which, upon requesting information, analyze the related labels and prevent the viewing of information which has not been authorized (according to the user's preferences).

---

<sup>3</sup> [www.rsac.org](http://www.rsac.org)

<sup>4</sup> [www.safesurf.com](http://www.safesurf.com)

There are three main methods of filtering:

### **Filtering by Browser**

The Microsoft Internet Explorer<sup>5</sup> Ver. 3.02 already supports filtering using PICS and it is anticipated that the Netscape Communicator will also support it in its forthcoming version.

In one of the option screens, it is possible to determine, based on the type of categories of different authorities, the threshold levels beyond which the browser will not allow to display information. This is an inexpensive solution (free of charge) and does not require special installation. It is though limited:

The filtering of the information is relating only to that that is brought by the browser, and there is no protection against information that arrives via IRC, FTP, e-mail or discussion groups.

The restrictions can be override by installing a different browser.

### **Designated Filtering Programs**

Comprehensive software package that monitors the information flow in and out of a computer. These programs enable control of other types of information as well, by filtering the computer's communications channels (accordance to the user's preferences).

The programs also support individualized definitions for the enabling access filtering to the different sources of information according to different user profiles.

These packages generally include additional safety devices, for example, CyberPatrol<sup>6</sup> (today's market leader) which enables the blocking of transmission of one's personal particulars (name, address, credit card number, etc.), even during the course of IRC conversation, definition of the permitted hours of activity per user, locking of local computer applications (accounting programs, for example) and protection against attempts to cancel or erase programs.

Software programs of this nature are the best solution for the home and private businesses.

### **Filtering by the Server**

Software programs similar to those mentioned above can be purchased as server versions. These programs connect to the proxy server and enable filtering for everyone using the network. This capability is very important for large organizations who wish to prevent improper use of the network during working hours, or to enable such use only during restricted hours (breaks, night shifts, etc.).

The advantage of such an installation is the ease of maintenance since it eliminates the

---

<sup>5</sup> [www.microsoft.com/ie](http://www.microsoft.com/ie)

<sup>6</sup> [www.cyberpatrol.com](http://www.cyberpatrol.com)

need to install and update versions in end-user stations.

Large organizations can also form their own independent evaluation authorities to sort and label, in similar manner to the general authorities, using categories and criteria which are suited to their organization. This makes it possible, by using standard, universal tools, to sectionalize information even within the organization. Another advantage in the use of filtering mechanisms of this type is the adding of additional defenses against viruses or other programs that are found on the Internet.

It should be remembered that the addition of programs is likely to somewhat slow the connection to the network because it serves as the bottleneck for the flow of information.

### **Additional Uses**

The motivation for the labeling mechanism was primarily designed to avoid accessing negative information but in the near future, it is expected that the use of labels (in PICS standard) will develop for other purposes in the fields of sorting and organizing information on the Internet.

This expectation arises from the fact that the more the information and the number of users of the Internet increase, the problem of searching for information on the Internet becomes one of qualitative location of relevant information.

Examples of uses on the horizon:

Qualitative assessments of sites and information on different subjects - similar to what exists at the moment in a number of catalogs. It will be possible to obtain in future, information about the quality of the site and the kind of information it contains. This will be done by organizations who create labels with this information without their dependence on specific sources for their information.

The addition of the information contained in the labels will enable the search engines to work with greater abilities. This ability will enable more efficient searches integrating the mechanized abilities for locating of information with human evaluation.

The use of labels for classifying users in discussion groups or messages posted from them. In addition to the filtering which is mentioned above, it will be possible to evaluate the message writers based on their messages. Filtering categories such as professional interest level, style, level of succinctness and more will enable the user to filter the messages by definition of thresholds.

And so on...

### **The Fly in the Ointment**

The process of labeling, and the filtering abilities described above constitute a degree of danger to two major principles of the Internet:

Censorship - in spite of the standardization of the labels, and intent to

establish a free market of different evaluators and classifiers, there are those who claim that a number of large organizations will control this field and establish a kind of elitism that controls the accessibility of the information on a global basis.

Popularity- resources required to carry out human labeling of the information and updating of this labeling, will mean that not every site could be labeled. It is expected that the use of the possibility to prohibit all unlabeled information from being accessed will create a situation wherein access to minority group sites will be barred. The same goes for non-institutionalized sites, whether this was the original intentional or not. This will affect the special character of the Internet, in which every site is accessible worldwide.

### **Summary - Filtering as a Service to Government Offices**

The major advantages of the Internet arise from the great accessibility of information in it to users worldwide. In our process of evaluation of this working tool, the advantage is also shown to be a disadvantage inasmuch as the existing organization of most institutions means that they are unable to monitor the use of this tool or to enforce regulations concerning its use.

Like all organizations, the various Government offices can use the filtering possibilities that have been described herein for the purpose of solving this problem, should it exist.

The Government Internet Committee decided neither to issue policy guidelines nor to enforce use of this nature. This is because the decision to use this tool and the method of filtering by it are issues that should be determined by the individual offices.

In the event that a Government office does decide to use the Internet in this way, we recommend that it should be done by purchase of specialized filtering software, which supports PICS, on the office server. Use of this nature will enable central regulation and bring about the greatest degree of efficiency.

### **Resources**

We recommend the sites motioned in the footers and in addition:

Article in the Scientific American by Prof. Paul Resnick, Chairman, PICS -

[www.sciam.com/0397issue/0397resnick.html](http://www.sciam.com/0397issue/0397resnick.html)

Article in Hot Wired, by Simson Garfinkel -

[www.hotwire.com/packet/garfinkel/nc\\_today.html](http://www.hotwire.com/packet/garfinkel/nc_today.html)

PICS homepage in W<sup>3</sup>C -

[www.w3.org/PICS](http://www.w3.org/PICS)

[www.w3.org/pub/WWW/PICS](http://www.w3.org/pub/WWW/PICS)

## Appendix A. - Technical Definitions

The technical definitions of the PICS standard are based on text with the intention being to make them easy to construct and update.

As we have already mentioned, the standard defines two elements:

The authority that classifies definition of vocabulary or categories based on which its labels are defined.

The label itself.

For demonstration purposes, we will characterize definitions and type based on the American standards for movie classification (MPAA).

### Classification Factor

In the URL address which defines the classification should be part of the page which includes the text:

#### The Text

((PICS-version 1.1)

(rating-system "http://MPAAAscale.org/Ratings/Description/")

(rating-service "http://MPAAAscale.org/v1.0")

(icon "icons/MPAAAscale.gif")

(name "The MPAA's Movie-rating Service")

(description "A rating service based on the MPAA's movie-rating scale")

(category

(transmit-as "r")

(name "Rating")

(label (name "G") (value 0) (icon "icons/G.gif"))

(label (name "PG") (value 1) (icon "icons/PG.gif"))

(label (name "PG-13") (value 2) (icon "icons/PG-13.gif"))

(label (name "R") (value 3) (icon "icons/R.gif"))

(label (name "NC-17") (value 4) (icon "icons/NC-17.gif"))

#### The Meaning

Site address of access

Organization site address

Organization Symbol

Organization Name

Organization Description

Definition of Category  
(word)

For each (numerical) level of the category the name is defined (or the literal description of the criteria) and a symbol which shows it for graphics purposes in the filtering programs

### The Labels

The following text can be included in the headers of a page of information or separately.

#### The Text

#### The Meaning

(PICS-1.1)	
"http://MPAAscale.org/v1.0"labels	Type of label (organization of filtering which will define it)
no "1996.11.05T08:15-0500"	Date it became valid
until "1997.12.31T23:59-0000"	Date it became invalid
for "http://www.company.com/filmname.htm"	information referred to (enables separate placement)
by "John Doe"	Evaluator
ratings (r 2)	Rating based upon each category (one in this case)

1.The World Wide Web Consortium (W3C for short) is a joint venture of industry and universities in the USA, Europe and Asia, for development of protocols

2.For details see <http://www.w3.org/pics> or <http://www.w3.org/pub/www/pics>

3.The program includes a list of recommended and forbidden sites, and thereby is acts as a labeling authority as well – <http://www.cyberpatrol.com>