

A man in a dark suit and white shirt is looking through a magnifying glass held over his right eye. The background is a blue gradient with faint binary code (0s and 1s) scattered around. The text "A Replica of Human Intelligence in Web Content Filtering!" is overlaid on the left side of the image.

A Replica of Human Intelligence in
Web Content Filtering!

NetProtector

A Neural Network Based Filtering Solution

an overview on a High Speed, Accurate, and dynamic content filtering solution

Abstract

Today's Web 2.0 genre of internet users is boundless to access anything under the sun and transcends all the borders of conventions and ethics. Uncontrolled and negligent usage of internet comes with a package of hazardous effects which can affect the growing children, slash the employee productivity, and distract students. These security and productivity ramifications compel the parents, educators and employers to enforce a stringent internet usage policy or employ an internet filtering solution. This paper discusses the need of internet filters and the difference that the NetProtector can make.

Introduction: New Challenges, New Opportunities

With the ever changing trends of technology, the web content filtering is facing new challenges and opportunities. In the early days of URL filtering, the challenge was to gather as many pre-categorized URLs as possible for the database. This would make it unlikely for a student or employee to view the objectionable Web pages. However, this technology suffers from the following drawbacks in today's world of pro-active pornographypredators:

- There is a great deal of subjectivity with human review.

- The content on many websites is constantly changing and a database of pre-categorized URLs cannot provide adequate filtering in such a dynamic Internet environment.
- URLs change very often and keeping up with these URLs is a challenge, as it requires the database to be updated frequently to maintain consistency and accuracy. Maintaining an accurate blacklist is extremely difficult and maintaining a categorized list of the entire World Wide Web is nearly impossible.

Millions of new URLs are created, intensifying the above challenges. With the highly dynamic nature of internet, Static URL Filtering is not effective anymore. Several players in the content filtering business have realized this gap and made good efforts to take the challenge. Some of them did fairly well and came up with effective dynamic filtering solutions to categorize the web pages at real-time.

The NetProtector Solution

NetProtector provides a host of features and filtering options that use the State-of-art technology to scan the content. NetProtector employs a Multi layered filtering technique which guarantees to block the unwanted content. Unlike the other internet filtering products with limited levels of filtering, NetProtector provides both static and real-time filtering. Starting with the preliminary database lookup and URL Analysis the content can be gauged up to the neural network based dynamic filtering.

Multi Layered Filter

NetProtector is an advanced web content filtering solution with multiple layers of filtering:

- **NetProtector database:** The evaluation of any URL request starts from the NetProtector database. The database, which is the Knowledge Base, contains a list of URLs distributed among grey, white and blacklist tables based on their nature of content. When the URL is not found in the database, the request is forwarded to the next layer, URL Processor.
- **URL Processor:** The URL Processor evaluates the terms in the URL of the requested Web page to determine its nature. The URL Processor takes weighted approach by maintaining a list of terms, where a value (weight) is assigned to each term. The value, which can be negative or positive, represents the degree of objection associated with that term. If the cumulative weight of the terms present in the URL exceed the defined threshold value, the Web page is blocked, else it is sent to the next level of filtering.
- **Internet Content Rating Evaluator (ICRE):** Most Web pages carry a rating in the metadata for the contents they display. The ICR

evaluator uses this rating to judge the Web page. If the internet content rating in the Web page declares its content to be objectionable, the ICR evaluator blocks the Web page. If the Web page does not have any rating, or if the rating is not objectionable, the ICR evaluator forwards it to the neural network for next level of filtering.

- **Neural Network:** The neural network has its roots in the science of artificial intelligence and is used to evaluate the contents of a Web page in real-time. This is the heart of NetProtector and it is unique. The content is scanned thoroughly by the Neural Network and then a decision to block or release the page is made. Neural Network is extensively trained for categorizing pornographic content. It can be trained for any category with a little effort. Knowledge Base is updated with the results of the filtering, so that the black white lists are up-to-date, thus eliminating the need of repeated real-time scans of the pages .

Dynamic Language Identification

NetProtector identifies the language of the web page content in real-time and employs appropriate language specific Neural Network for classification.

Multi lingual Neural Networks

Neural Networks are extensively trained for English, Dutch and French languages and can be easily trained for other languages.

Safe Search

NetProtector automatically turns on the Safe Search option in all the popular search engines and forbids the access to objectionable content at the initial stage itself.

Chat Filtering

NetProtector filters inbound and outbound chat transcripts. It prevents kids from being exposed to inappropriate content within chat sessions.

Easy Administration

The easy-to-use interface helps you

- Track the issues reported by users.
If NetProtector accidentally blocks or allows certain websites, you can report this erroneous behavior to the Administrator. Using this interface, an Administrator can easily monitor these issues and provide appropriate resolution.
- Update the database with URLs and offensive terms in an effortless manner.

Tailor-made solutions for all users

Internet filtering needs vary for different users. Be it entrepreneurship or education or for a family, NetProtector allows you to customize the filtering solution as per your convenience.

The Parental Control feature enables you to define the sites that can be accessed by the family members and limit the timings and time period of the internet access.

The Server-based solution is found highly useful for corporations, schools, and ISPs. With the server-based solution, you can leverage the filtering capabilities to all the computers connected to the network without having to install individually.

The Provider Independent Client Solution is most suited for the home users. You can continue with your existing Internet Service Provider and leverage the offerings of the NetProtector. Client solution is similar to a plug-in and is installed on the client's machine. It in turn contacts a remote NetProtector server for filtering instructions. All the web filtering requests are answered quickly and the latency time is negligible. This is because the NetProtector Server sends only the block/allow signal to the client solution and the website content is retrieved from your local ISP.

High Speed, Accuracy and Scalability

NetProtector is a one-stop-shop solution for all the internet filtering challenges. It can simulate human smartness to judge the content of the page at a very high speed. In this process, the latency goes unnoticed as it is very negligible. NetProtector offers multi level filtering such that results are accurate. We ran an analysis of our customers and realized that they face a 0.0003% of erroneous blocking on an average day with 23027294 total hits and a maximum of 70 reported cases. Considering the unidentified cases, we claim more than 98% accuracy in filtering pornographic content.

NetProtector is highly scalable to suit the needs of all varieties of businesses. From a small school/organization to a large ISP, NetProtector can be deployed anywhere without compromising on reliability. Our Customers vary from a 300 employee organization to an ISP with more than 50,000 subscribers.

Summary

The nature of Web traffic and browsing habits has evolved far beyond the early generation URL filtering architectures. Enforcing appropriate use policy and providing robust Web content security requires a truly dynamic filtering solution. NetProtector uses a combination of leading- edge dynamic techniques in its architecture to provide the most accurate and most effective content filtering solution.

NetProtector is the only solution that provides a comprehensive, best-of-breed, and integrated functionality at a low TCO.