

# Dynamic Content Analysis & Endpoint Content Filtering Overview

## *“The Facts and FAQs”*

Content Filtering providers offer diverse ways to filter and categorize Internet content. The question of whether it's better to (a) *filter Internet content based on a list of URLs* or (b) *use a “smarter” filtering engine to dynamically analyze every website users visit and categorize this content in real-time* is an important topic as more and more children become daily smart device users.

This document discusses the advantages and evaluation methods that can be used to validate filtering techniques and discusses the benefits of End Point Content Filtering vs. Cloud or other web-based services.

### **Filtering Techniques & Why Dynamic Content Analysis is Important**

Generally, Web Content Filtering is accomplished via one of two base methods – List Based Filtering or Dynamic Filtering.

**List based filtering** is based on a catalogue of websites or domain names with an associated category for filtering. Every time a website is requested, a service validates the URL or domain against this database to verify that the user or network settings permit the URL's associated category. Many vendors that utilize this technique use the size of the URL database and frequency at which it gets updated as selling points in product comparisons.

**Dynamic filtering** analyzes the web content and looks for category specific content to identify the category or type of content found within the web page rather than of referencing a catalogue of websites.

Within the general method of Dynamic filtering, the complexity of intelligence used to determine web content types varies widely. Although these methods are not detailed in this synopsis, it is important to understand that all Dynamic filters are equally effective or accurate.

In comparing Internet Filtering techniques, it's important to mention that list-based URL filters have an implied limitation. This limitation is caused by the web's constantly changing content. The name “list” implies that a static URL list is created, but because of the changing nature of the Internet, this list is likely out of date as soon as it's published. To offset this issue most URL list vendors, dedicate teams of people to manually categorize new websites. Unfortunately, thousands of new websites are posted every day and most websites have content that is frequently changing or contains user generated content, making it impossible to keep up with the required changes to a list.

ContentWatch products utilize Dynamic Contextual Analysis (DCA) to filter and categorize web content. This technology provides up-to-date and accurate filtering. A DCA engine analyzes the content of every web page in real time and uses custom algorithms to categorize or block this content based on the settings or selections of blocked categories. ContentWatch has been enhancing this dynamic engine to become efficient and accurate for more than 10 years and its proprietary technology will block or allow a user to view a web page based on its content in real time.

(801) 508-3600

Fax (801) 508-3595

6949 S High Tech Drive  
Salt Lake City, Utah 84047

## Frequently Asked Questions about Dynamic Contextual Analysis (DCA)

### 1. How does Dynamic Contextual Analysis (DCA) work?

DCA reassembles each requested web page and then scans the page for specific content types. This is done by searching through the text and format of the page. While looking at this text the DCA engine evaluates key words and phrases and the context in which these words and phrases are being used. Other key indicators that are also evaluated are links, metadata, header information, site ratings and names of videos and images. With this proprietary technology, the analysis engine can instantly determine the web content type on each web page.

The DCA engine determines what types of content are found on each web page in milliseconds and then blocks the pages with content and categories that are set to “block” by the software settings.

DCA is smart and intelligently considers misspelled words and slang speech within each web page. DCA can also analyze URLs, RTA (Restricted to Adults) and PICS (Platform for Internet Content) ratings, embedded URLs, page frames and other metadata to ensure a more accurate categorization.

DCA also uses context in order to more accurately categorize content that may not be easily classified. DCA can differentiate whether the word “breast,” for example, is being used in a medical context (such as *breast cancer*) or in a cooking context (such as *chicken breast*) or in a sexual or adult/mature context (such as *breast augmentation*), and then will block or allow that content based on the software’s policy settings.

### 2. While evaluating a product using DCA I was unexpectedly blocked on a website.

DCA does not look for “themes” of websites, it only looks for categories of content found within each web page. This means a website that typically has safe or “known” content types could have a blocked page if the content on the page changes or includes content categorized to match a blocked category.

An example of this might be found on news site such as [www.cnn.com](http://www.cnn.com) or [www.msnbc.com](http://www.msnbc.com). If DCA is set to detect and block violence and a news story is focused on violence and contains violent content, that page could be blocked.

While evaluating the accuracy of DCA, check the block message for information about which category of content was detected and then verify that you want that type of content blocked. Typically, if you return to the page without the DCA software running you will see that this web page in fact contains the blocked content type.

Since DCA is intelligent, blocking a category such as “gambling”, will not just block gambling websites, but will also block any website that contains “gambling” content.

### 3. How can I see if DCA is working or accurate?

To best see how DCA works and verify its accuracy we recommend that you install ContentWatch on a device you wish to use to test, then configure the solution in a restricted setting that blocks many categories of web content and following some of the suggested evaluation methods contained in the final pages of this document.

### **Summary of Dynamic Content Analysis (DCA):**

- *More accurate and faster than list base solutions.*
- *Real-time analysis on every web page making it more comprehensive and complete than List Based Filtering.*
- *Searches and determines categories of content through intelligent and contextual algorithms.*
- *Blocks any web page that contains unwanted content – not just entire websites or domains.*
- *Analyzes all content and searches to ensure that protection is always available.*
- *Customizable and improves usability and accuracy.*

### **Local Endpoint Filtering vs. Cloud or Network Services**

Beyond the methodologies used to filter content and determine content types, an equally important topic is the location where this deterministic technology resides and is implemented. In simple terms, the question of “where” does the filtering occur needs to be addressed.

Before the Internet became highly accessible and high speed bandwidth connections were readily available in every home, office and via cellular providers, endpoint solutions were considered the only viable option for Content Filtering.

Along with slow and unstable Internet, older computers suffered from less computing power and limited hardware resources. This limitation often resulted in slow computer performance when additional processing power was needed to analyze Internet content.

As the Internet became more reliable and Cloud or Network Services became adopted, many content filtering providers moved their services to the Cloud. At first glance, this seemed like an acceptable solution, however, it introduced new problems associated with performance and filtering effectiveness.

Cloud or Network solutions for Content Filtering exposed the following deficiencies:

#### **1. Added latency for network traffic**

The added latency to check with a remote service or routing traffic to a remote service via VPN or Proxy-like solutions attributed to slow Internet surfing and poor user experiences.

#### **2. Dependency on remote services**

Although cloud introduces redundancy and is typically performant, the reliance on a remote service for Internet activity can introduce a new single point of failure.

#### **3. Issues with encrypted network data**

With common encryption standards, remote services are ineffective at analyzing HTTPS/SSL web traffic. This results in reverting back to list based solutions or just being ineffective at filtering secure web content.

#### **4. Additional bandwidth consumption**

Vendors and consumers incur additional expenses as bandwidth costs are incurred for the clients being filtered and the service hosting the filtering.

While other vendors have focused their efforts on moving their services to the cloud, ContentWatch has focused on optimizing its Content Filtering Engine to run locally. This engine has a very small footprint and has been optimized eliminate any used perceived slowness or latency.

The ContentWatch Dynamic Content Analysis Engine can analyze web content in real-time for secure and non-secure web traffic and be completely transparent to the user. This engine was originally optimized to run on computers with low CPU capability so modern machines with greater than 1GHz processors have no performance impact at all and users experience no additional latency or slow done.

In addition to no latency, local services for Content Filtering are more secure and can accurately analyze HTTPS/SSL secure web content. Since the solution runs locally, it is not dependent on a specific network connection or remote service. These attributes ensure that the solution provides protection on all web content and in any location.

With thousands of users, Network or Cloud based filtering solutions have limited customization options because the filtering service uses precious resources to track individual use and apply user specific settings.

Distributing a local filtering agent to each endpoint device allows the solution to be fully customizable and harness the available CPU and resources of each device. This system provides the best user experience and most comprehensive protection.

Current technologies including simple smartphones and tablets are not deficient in processing power and the overhead of local analysis is much less than the latency introduced by adding Network or Cloud services. This clearly demonstrates the advantages of a local filtering engine and agent on each endpoint device.

### ***Summary of Local Content Filtering and its advantages:***

- *Most performant while harnessing the available power of the local CPU instead of waiting for remote Cloud or Network services.*
- *No latency with Internet traffic is introduced so the user experience does not perceive any slowdown.*
- *Best solution for filtering and monitoring secure HTTPS/SSL traffic while remaining transparent and compatible with encryption standards.*
- *Most versatile for customization and configuration options since the logic is applied locally.*
- *Most cost effective for the provider and user, since no additional data or bandwidth consumption is required. DCA looks for and determines categories of content through intelligent and contextual algorithms.*

**ContentWatch** has created an exceptional hybrid solution with a **Hosted Cloud Service** for administration, configuration, reporting and monitoring and a **Local Device Application** for the Content Filtering Engine and deterministic logic. This solution gives Administrators and Parents the modern benefits of remote and cloud administration tools and also users the optimal user experience from real-time content analysis locally protecting and monitoring their device.

(801) 508-3600

Fax (801) 508-3595

6949 S High Tech Drive  
Salt Lake City, Utah 84047

### In Conclusion

Content Filtering has evolved in both filtering techniques and its method of implementation. To provide the best user experience and most accurate Content Filtering experience a solution must include robust and optimized Dynamic Content Filtering technology. This technology must be real-time and intelligent to keep up with constantly changing content.

Filtering technology is best applied locally on each endpoint device where the solution can be fully customized and use the available hardware resources to analyze, monitor and filter the content. Implementing locally not only provides the best performance, but also provides the best options to prevent circumvention and filter secure web content.

ContentWatch has the most intelligent and accurate Dynamic Content Analysis engine available today. This solution has been enhanced for accuracy and performance for more than a decade. ContentWatch has also created the endpoint software for all major platforms to implement this filtering locally on each device, where it is most effective and provides the best user experience.

To learn more about Dynamic Content Analysis and ContentWatch technologies that implement this technology please visit [www.contentwatch.com](http://www.contentwatch.com).

### ContentWatch is Trusted by:



STANFORD  
UNIVERSITY



Sprint™

verizon✓



KPMG



SaskTel



Knowledge Learning Corporation®

(801) 508-3600

Fax (801) 508-3595

6949 S High Tech Drive  
Salt Lake City, Utah 84047

# DCA Evaluation Methods

## **Method 1: Browsing websites with mixed or often changing content.**

The benefits of DCA are best recognized when accessing websites with mixed content types where one page may contain content that is appropriate, but other pages on the same site may not. Below are some mild yet effective examples of DCA working and effectively categorizing content.

While evaluating the links below, please note that some web pages' load and others are blocked because a specific content type was found on that page. The categorization is not determined by a single keyword or by pre-determined list, but by the actual context and content on that web page.

### **Walmart ([www.walmart.com](http://www.walmart.com))**

URLs that aren't block for any categories:

- <http://www.walmart.com/cp/televisions-video/1060825>
- <http://www.walmart.com/store/1686/weekly-ads>
- <http://www.walmart.com/store/finder>

URLs that DCA should detect and block based on associated content categories:

- <http://www.walmart.com/ip/Benjamin-Trail-NP2-with-Scope-.22-Caliber-Air-Rifle-Camo/37001625> (Weapons)
- <http://www.walmart.com/cp/Intimates-Loungewear/1078024> (Lingerie/Swimwear)
- <http://www.walmart.com/ip/Trojan-Sensitivity-Ultra-Thin-Premium-Lubricant-Condoms-36ct/20896076> (Mature)

### **Pinterest ([www.pinterest.com](http://www.pinterest.com))**

Pinterest is generally not considered a 'bad' site. However, many categories of content exist within the site and there are pages that contain content that match categories that an administrator may choose to block.

URLs that aren't blocked:

- <https://www.pinterest.com/pin/55380270394268565/>
- <https://www.pinterest.com/pin/1266706122743771/>

URLs that are blocked:

- <https://www.pinterest.com/pin/17662623508450653/> (Nudity, Provocative)
- <https://www.pinterest.com/pin/239394536419061315/> (Profanity, Abortion)
- <https://www.pinterest.com/pin/374713631469926548/> (Profanity, Gambling)

### **Wikipedia ([en.wikipedia.org](http://en.wikipedia.org))**

DCA allows for each page and its content to be categorized in real-time and within each page. While on Wikipedia, try a sample search for "Chicken Breast" or "Chicken Breast Recipes", notice that these pages load correctly. Then try searching for "Best Breasts" and notice that it is blocked and categorized as "Nudity and Adult Mature".

Wikipedia examples include:

- <https://en.wikipedia.org/wiki/Alcohol> (wiki page on alcohol - chemistry related) Allowed
- <https://en.wikipedia.org/wiki/Beer> (wiki page on beer) Blocked for Alcohol
- <https://en.wikipedia.org/wiki/Weed> (wiki page on garden weeds) Allowed

- <https://en.wikipedia.org/wiki/Marijuana> (wiki page on marijuana) Blocked for Drugs
- [http://en.wikipedia.org/wiki/Colt\\_45](http://en.wikipedia.org/wiki/Colt_45) (this leads to an index page with no questionable content)
- [http://en.wikipedia.org/wiki/Colt\\_45\\_Single\\_Action\\_Army](http://en.wikipedia.org/wiki/Colt_45_Single_Action_Army) (Weapons)
- [https://en.wikipedia.org/wiki/Radiocarbon\\_dating](https://en.wikipedia.org/wiki/Radiocarbon_dating) (wiki page on carbon dating) Allowed

### **Olive Garden ([www.olivegarden.com](http://www.olivegarden.com))**

Pages on this site generally won't be blocked, but if you navigate to the alcoholic beverages and have the category of Alcohol blocked, DCA will see the content and block it.

URLs that aren't blocked:

- <http://www.olivegarden.com/home>
- <http://www.olivegarden.com/menu-listing/dinner>

URLs that are blocked:

- <http://www.olivegarden.com/menu-listing/wines> (Alcohol)
- <http://www.olivegarden.com/menu-listing/cocktails-beer> (Alcohol)

### **Method 2: Browsing to common sites that should be blocked**

Commonly known and categorized websites are easily detected and blocked based on their content types. These results are very similar to what would be found in a list based solution because these sites are well known and keep a constant or consistent type of content across all web pages within their site.

- [www.hustler.com](http://www.hustler.com), [www.redtube.com](http://www.redtube.com) and [www.playboy.com](http://www.playboy.com) are easily detected as pornography.
- [www.onlinecasino.com](http://www.onlinecasino.com), [www.place-your-bets.com](http://www.place-your-bets.com) and [www.gambling.net](http://www.gambling.net) are blocked for gambling.
- [www.victoriassecret.com](http://www.victoriassecret.com) and [www.maxim.com](http://www.maxim.com) are blocked for provocative and lingerie.

### **Method 3: Searching for content that should be blocked**

DCA can detect returned search results and block inappropriate content. Often times, search results do not contain sufficient content to determine a category and in this case DCA will let the search page load, but will block the web page from loading when clicking on a link.

#### Example:

Search for beach and the results will return with websites and pictures of beaches. If you search for "beach bodies", the results are blocked and categorized as provocative.

Additionally, DCA will enforce search engine "Safe Search" technology to all searches. This will help DCA to detect and block "image only" searches found within search engines to protect against inappropriate images.

### **Method 4: What happens if DCA incorrectly categorized a web page?**

DCA is the most accurate and effective filtering technology available today, but just like any technology it is not perfect. To help users have the best experience possible, DCA will allow users to re-categorize or change the category of any website or web page. This provides the best experience for users that need access to content that may be incorrectly

categorized without having to wait for a list based solution to manually review and add a website to a list.

To improve DCA, re-categorized web content is evaluated and used to better train the DCA filtering engine to constantly improve accuracy and effectiveness. DCA uses a simple definition (DAT) file as a filtering ruleset that can be transparently updated to ensure all users have the best filtering accuracy and technology available.

###

End

(801) 508-3600

Fax (801) 508-3595

6949 S High Tech Drive  
Salt Lake City, Utah 84047