A European Prototype NETPROTECT of Internet Access Filtering

EADS MATRA SYSTEMES & INFORMATION

Net Protect

OPTENET.com OPTIMAL INTERNET

matra global netservices

HYPERTECH INFORMATICS & INTERNET SOLUTIONS

SAILLABS

D2.3 - Report on Filtering Techniques and Approaches

version 1.0

Contract NETPROTECT 26511

Report on Filtering Techniques and Approaches

Project	NETPROTECT	**Contract**	NETPROTECT 26511
Client	European Commission		
Reference	NETPROTECT:WP2:D2.3		
Issue (draft)	1.0	**Date**	23 October 2001
Status	Definitive	**Nature**	Public
Author(s)	Ana Luisa ROTTA	**Organisation**	Red Educativa

	Name	**Role**	**Organisation**
Checked by	Ana Luisa ROTTA	Deliverable Leader	Red Educativa
Approved by	Sylvie BRUNESSAUX	Work Package Manager	MATRA Systèmes & Information
Authorised by	Stéphan BRUNESSAUX	Project Director	MATRA Systèmes & Information

Distribution list	
Name	**Organisation**
Gerhard HEINE	European Commission
Stéphan BRUNESSAUX	EADS Matra Systèmes & Information
Francisco MARTIN	Red Educativa
Olivier ISIDORO	Matra Global Netservices
George FERLIAS	Hypertech
Johannes RITZKE	Sail Labs

Synopsis

This document provides the report on the different techniques and approaches used in filtering at the present time.

File name

OPT-WP2-D2.3-v1.0.doc

Amendment History

Version	Date	Description
0.1	26 June 2001	First template.
0.2	07 August 2001	Second template
0.3	20 August 2001	Third template
0.4	22 August 2001	Fourth template with conclusions and executive summary
0.5	24 August 2001	Final draft handed to Project Manger
0.6	03 September 2001	Review of first draft
0.7	21 September 2001	Comments made on review record addressed
0.8	18 October 2001	Last comments made by Sail were added
1.0	23 October 2001	Final editing.

Scope of NetProtect

NetProtect is a project partially funded by the European Commission under the Safer Internet Action Plan and related to the development of rating and filtering systems for Internet content.

The objective is to build a European prototype of an Internet access filtering tool for parents and teachers which addresses the problems of current existing filtering solutions: inappropriate blocking/filtering techniques which sometimes blocks legitimate Web sites and occasionally allow questionable Web sites, inability to filter non-English Web sites and therefore most Web European sites, lack of transparency disabling the user the right to know why some sites can be accessed and not others.

The project is carried out by a consortium led by EADS Matra Systèmes & Information (France) and grouping together Red Educativa (Spain), Matra Global Netservices (France), Hytertech (Greece) and Sail Labs (Germany).

The project officially started on 1 January 2001. It will end on 1 May 2002. The overall budget of the project is 800 622 Euro and the total effort is 7.1 man-years.

Executive summary

This report presents the different filtering techniques and approaches used by various tools, products and prototypes. The basic idea of this report is to analyse the maturity of the various filtering techniques and approaches that can be used for developing Internet filtering solutions.

For that, the different techniques used for the development of Internet filtering solution were divided into two major groups:

- Text-based content analysis techniques;
- And, image analysis techniques.

The first group included all of these techniques and approaches that use text based content analysis. The second group includes all of the solutions that classify the type of web page content, based on the analysis of the images contained in the page.

Furthermore, the consortium located various tools, products or prototypes using text-based content analysis technique and tools, products or prototypes using image analysis techniques, as well as a few tools, systems and prototypes that use both image and text-based content analysis techniques. A review of these tools, products or prototypes was conducted and the criteria for their analysis were established.

It was seen in the analysis that for the tools, products and prototypes that use only text-based content analysis techniques the results obtained are very similar to the results of those systems that use only positive and negative lists. Nevertheless, it seems that the combination of text-based content analysis techniques with positive and negative lists might improve considerably the performance of filtering solutions.

Looking at the issue of image analysis techniques combined with text-based content analysis techniques, again the results are not better than those obtained using COTS based on positive and negative lists.

Regarding the systems that only carry out image analysis techniques as it is the case one might conclude that although these systems are able presently to identify skin with very high degrees of accuracy, they are not able to distinguish between what is just skin and what is nudity and much less what is pornography. However, certain filtering solutions that use image analysis techniques can really bring an added value to an integrated solution as the one we intend to develop in the context of the NetProtect project

Therefore, one may conclude that text-based content analysis and image analysis, at the present moment with these technologies as they currently stand, none of them present better results than those systems that work exclusively with the use of positive and negative lists.

However, it seems that a good way to improve the systems available presently and to reach the objective of NETPROTECT, i.e. building a European prototype of an Internet access filtering tool for parents and teachers which addresses the problems of current existing filtering solutions seems to be the combination of the different techniques with the system of lists. It appears that, at the present moment, this is the only viable solution in order to address the problems that might arise with currently available filtering technologies. With the combination of different techniques in filtering the following problems could be more easily addressed:

- Inappropriate blocking/filtering techniques which sometimes blocks legitimate Web sites and occasionally allow questionable web sites;
- Inability to filter non-English Web sites and therefore most European web sites.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The objective of NetProtect is to build a prototype of a multi-lingual Internet access filtering tool for parents and teachers that addressed the problems parents and teachers might have with the current existing filtering solutions:

- Inappropriate blocking/filtering techniques which, in some occasions, block harmless web sites and allow access to inappropriate web sites.
- The inability to filter non-English web sites reducing their utility for most European languages.
- Lack of transparency denying the user the right to have access to information regarding the reason for the blocking or not of web sites.

The prototype tool is intended at preventing the access of children to pornographic web sites regardless of their language (English, French, German, Spanish or Greek). It will be developed by integrating existing COTS tools, such as filtering products based on red/green lists, language products, text analysers, image analysers, etc.).

The consortium expects to extract the following benefits from the project:

- A NetProtect web site providing European Internet users with useful information on filtering solutions;
- A review of existing filtering tools;
- A review of existing techniques and approaches that might be adapted to filtering;
- An analysis of European users requirements in terms of filtering solutions;
- A prototype of a multi-lingual Internet filtering software able to filter pornographic web sites in. English, French, German, Spanish or Greek.
- A full evaluation of the prototype;
- Recommendations for a full-scale industrial product.

This report presents the different filtering techniques and approaches used by various COTS filtering tools. The basic idea of this report is to analyse the maturity of the various filtering techniques and approaches that can be used for developing Internet filtering solutions. Specific attention is again given to the evaluation of these techniques that deal with web sites in English, German, French, Spanish and Greek.

The different techniques used for the development of Internet filtering solution were divided into two major groups:

- Text-based content analysis techniques;
- And, image analysis techniques.

The first group includes all of these techniques and approaches that use text based content analysis. The range of solutions within this group is rather wide comprising from products which blocks access when certain words appear in the page textual content, to techniques which carry out semantic analysis of the text and content recognition.

Image analysis techniques include all of the solutions that classify the type of web page content, based on the analysis of the images contained in the page.

The consortium located various tools, systems or prototypes using text-based content analysis technique and tools, systems or prototypes using image analysis techniques, as well as a few tools, systems and prototypes that use both image and text-based content analysis techniques. A review of these tools, products or prototypes was conducted and the criteria for their analysis were established.

The original idea was to measure the effectiveness of these tools, systems or prototypes applying the same list of URLs used in the D2.2 - Report on currently available COTS filtering solutions [1], in order to evaluate the effectiveness rate of the identified techniques through their existing implementations. However, as it will be explained further in the report, this was not possible for certain tools, products and prototypes and their analysis was carried out differently. In addition, it is not always possible to determine the exact technique used by the different products, tools or prototypes, for this reason we have classified the tools, products and prototypes analysed according to the two main groups of techniques mentioned previously, without further classification.

Chapter two of this report presents an explanation of text-based content analysis techniques or text classification. The three types of text classification techniques, manual text classification, semi-automatic text classification and automatic text classification, are examined and special attention is given to automatic text classification. Furthermore, chapter 3 of the report presents an explanation on image analysis techniques or image filtering.

Chapters 4, 5 and 6 of the report provide a list of tools, products and prototypes identified by the group that use either text-based content analysis techniques or image analysis techniques or both. A brief description of these tools, products or prototypes is provided including information of their web page, the operating systems supported by them, their effectiveness rate and where to find a demo version.

Chapter 7 presents the testing of the different tools, products or prototypes: the methodology used in each of the test and the evaluation criteria. The results of the testing are presented on this same chapter. For those tools, systems and prototypes that did not allow an evaluation, a thorough explanation of their technique is provided.

2 TEXT-BASED CONTENT ANALYSIS TECHNIQUES OR TEXT CLASSIFICATION

Text classification is known under a number of synonyms such as document / text categorisation, routing or filtering and topic identification.

Basically text classification can be defined as *content-based assignment* of one or more *predefined topics (categories)* to texts. Text classification can be used for filtering and routing of text for topic-specific processing mechanisms or directly for humans. Applications are e.g. filtering of news articles for knowledge workers, routing of customer email in a customer service department, or detection and identification of criminal activities for police, military, or secrete service environments.

There are three types of text classification techniques as of today: manual text classification, semi-automatic text classification, and automatic text classification.

2.1 MANUAL TEXT CLASSIFICATION

Manual text classification is an extremely time consuming and expensive task and therefore cannot scale with the exponential growth of the Web.

List-based filtering techniques face a series of problems such as the explosive growth of the Web (over 2,000,000 new sites added each year). Furthermore, web sites are also removed at a substantial rate. Therefore, it is very difficult to maintain lists of harmful sites updated.

2.2 SEMI-AUTOMATIC TEXT CLASSIFICATION

In semi-automatic text classification human experts develop 'programs' that recognise the topics of interest. However semi-automatic text classification is very knowledge-intensive: human experts collect thousands of documents that represent typical examples for the topic; from these documents, individual words, collections of words, and co-locations of words are manually analysed by human experts and then transformed into rules and patterns. These rules and patterns can be applied in order to classify new text automatically.

2.3 AUTOMATIC TEXT CLASSIFICATION

In (fully) automated text classification, the 'programs' that recognise topics of interest are developed / configured by an automatic process. Two phases can be identified in automatic text classification, the *learning phase* and the subsequent *classification* or *application phase*.

2.3.1 THE LEARNING PHASE

In the *learning phase* users define topics by giving *sample documents (training examples)* for each of these categories. Most methods for automatic text classification also require *counterexamples* for each category that is sample documents that do not deal with the respective topic. In a standard application for automatic document classification such as news filtering, users assign categories to the documents of a selected collection by hand. Documents may be assigned to more than one category if they deal with several of the topics or if one has a hierarchy of topics. The document collection used for learning should be as representative as possible for the documents that one expects to encounter in the future. All documents that are not assigned to a category may serve as counterexamples for this category.

The topic learner component analyses the sample documents and identifies the content that is specific for each topic. For this analysis a combination of *linguistic techniques* and *statistical machine learning techniques* are applied. This analysis is essentially an *inductive reasoning* process that involves *generalisation* and *abstraction*. The output is a *model* for each topic which is represented by a set of *classifier parameters*. Most classifiers make *a priori assumptions* about the underlying model and its *complexity*. If the assumed model complexity is too high, *overfitting* can occur. This means that the model overspecialises with respect to the training examples and generalization to new previously unseen examples becomes bad. Overfitting is one of the biggest problems in automatic text classification and in machine learning in general. However it can be avoided by special techniques or by giving a lot of training documents.

Normally one assumes that the learning phase is invoked very rarely. Therefore high computational costs in the learning phase are acceptable if they lead to good text classifiers.

2.3.2 CLASSIFICATION PHASE

In the *classification phase* new (previously unseen) documents can be given to the topic classifier which returns a topic association (a rating or classification for each topic). Basically, it tells the user whether the document deals with the training topics and especially with which topics it deals.

It is assumed that a lot of documents have to be classified. In a news filtering application new messages constantly arrive and have to be filtered or routed. Similarly, an internet filtering tool has to be very fast. Therefore, efficiency is very important for the classification phase.

2.3.3 STATE OF THE ART

During the last years, the field of automatic text classification has evolved from academics into real world applications. A variety of systems exist which use / combine methods from the fields of statistics, information theory, machine learning, and neural networks. Most automatic text classification systems comprise three steps (in the learning and classification phase): *feature construction, feature selection, and the actual classification step (pattern recognition).*

In the feature construction step vector-representations (vectors of real numbers) of documents are generated. This step is necessary since statistical techniques and machine learning techniques can only work with numbers. Feature construction methods generally differ in the amount of linguistic and statistic sophistication that is applied. Usually simple word or letter n-gram features are used. So far it has not clearly been demonstrated that linguistic preprocessing pays off (however multiword and proximity information is generally considered as important). A *feature vector-representation* for a document is simply a vector of weights for all the features. These weights are based on the frequencies of the features in the document (and maybe also on their frequencies in a whole corpus).

Most feature construction methods produce hundreds of thousands or even millions of features (all features detected in the whole training corpus). Most classification methods cannot handle such *high-dimensional input* (computational costs for learning and/or classification become intractable). A further problem is that the model complexity for many classifiers increases with the dimension of their input. This means that high-dimensional input vectors can cause overfitting. Therefore, *dimensionality reduction* or *feature subset selection* methods have to be applied in order to identify the features that are important (characteristic) for the topic(s) of interest. There is no agreement on the best method. Widely applied are various versions of the *mutual information measure* (for feature selection) and *principal component analysis* (for dimensionality reduction).

In the actual classification step the reduced (shorter) feature-vector representations of the documents are used as input for the actual classifier. The classifiers used in automatic text classification systems range from very simple ones like K-nearest Neighbors, Rocchio´s Centroid, and Perceptron to more sophisticated ones such as Multilayer Perceptron, Decision Tree, and Support Vector Machine. These methods differ in their expressive power (the kind of functions they can compute and learn) and in their proneness to overfitting. Some recent studies (including [2]) clearly indicate that Support Vector Machines are superior to other classification methods because they avoid overfitting very successfully.

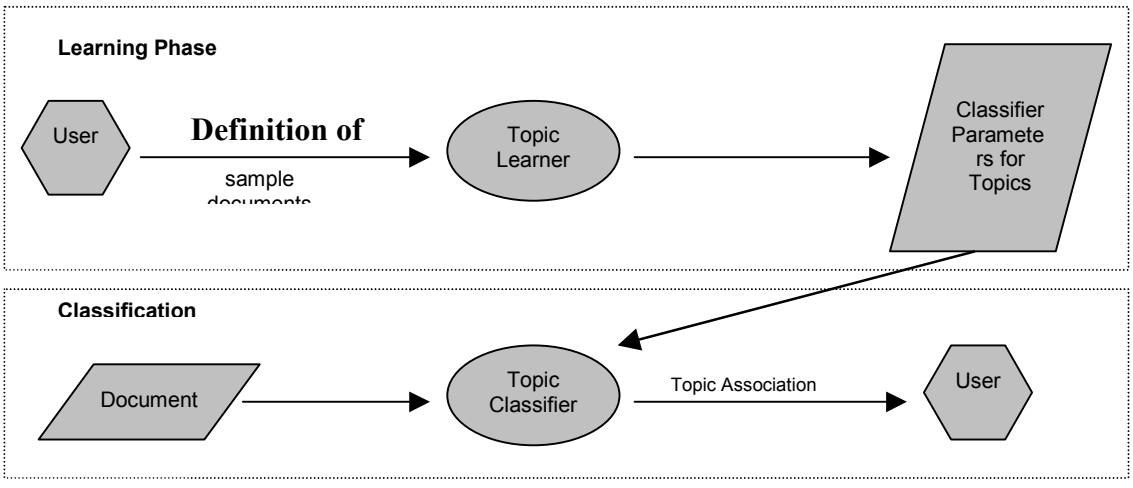


Figure 1: Document classification

3 IMAGE ANALYSIS TECHNIQUES OR IMAGE FILTERING

The approach used when analysing images consists in extracting generic features such as colour, texture, shape, etc. Colours are often represented by a histogram of colours that can be compared with other histograms. It is often the case that colour features are the basis for image characterisation. Using the techniques able to extract generic features such as colour, texture and shape it is relatively easy to identify skin, although it is sometimes easily confused with other things. Usually the skin is recognised due to the fact that it practically does not have texture and it is composed of a group of colour containing the red for the blood and the yellow and brown for melamine.

However filtering all images which contain a set of properties for colour, texture and shape of skin is too limited an approach for image filtering. In that particular case, the image analysis should provide a semantics level higher than just an indication of colour, texture or shape.

Some approaches to porn image filtering rely on an analysis of colour and texture with comparison with porn images stored on a database. Usually, the procedure consists first in a normalisation step consisting in converting the image in a format that can be exploited for the extraction of the image features. A histogram calculation is then performed in order to determine the skin colour percentage. A fine texture analysis is then carried out in other to determine specific descriptors. These descriptors are then compared to the descriptors of the images contained in the database in order to filter or not the image analysed.

Other approaches attempt to group the different parts identified as skin in order to see whether these parts are components of a naked body.

4 TESTING OF THE DIFFERENT TOOLS, PRODUCTS AND PROTOTYPES

The consortium's original idea was to measure the effectiveness of the tools, products and prototypes identified by applying the same list of URLs used in the D2.2 - Report on currently available COTS filtering solutions [1]. However, this was not possible for certain tools, products or prototypes and their analysis was carried out differently.

In order to carry out the review the consortium established a series of criteria in order to make the testing as uniform as possible for all of those tools, products and prototypes that allowed testing with CheckProtect. These criteria will be explained in the following sections.

4.1 CHECKPROTECT

As it was explained in D2.2 – Report on currently available COTS filtering tools [1], for each of the 5000 URLs, an application developed by the consortium called CheckProtect, calls Internet Explorer asking it to load each URL from the list of URLs. If the filtering solution decides to block this URL, it displays a message in the Internet Explorer window. CheckProtect detects this message and the URL is marked as blocked.

The results are stored in a file that can be processed to calculate the efficiency and the error rate of the filtering solution.

4.2 SQUID

To reduce the effect of the changing of the web sites, a proxy server (SQUID) is used. It stores the web pages in its cache and tries to serve the user with these pages when possible.

4.3 EVALUATION CRITERIA

This chapter presents the criteria suggested to carry out the evaluation of the tools, products or prototypes that use the different techniques and approaches identified by the group in the most objective way possible. The main objective of this chapter was to evaluate the different tools, products or prototypes in order to extract conclusions on the techniques used by each of them, i.e. to evaluate the different techniques applied in filtering through the existing implementations (tools, products or prototypes).

The first section of the chapter describes the categories in which the evaluation criteria have been grouped. In addition, there is a reference to the environment in which the study should be conducted in order to safeguard the neutrality of the study. Finally, the last section presents in a tabulated form each of the different categories of criteria to be measured.

It was not always possible to evaluate the different tools, products or prototypes that use the techniques selected by the consortium according to the criteria established. In addition, it was not always possible to determine with precision the exact technique used by the different tools, products or prototypes within the group of techniques established by the consortium. Nevertheless, as much information as possible has been collected according with the criteria presented.

4.3.1 CATEGORIES OF CRITERIA

Effectiveness – Measuring the effectiveness of a particular technology means analysing the performance of that technology in blocking pornographic content. In other words, this criterion will attempt to measure to what extent is the technology successful in preventing pornographic content from being displayed within a browser during an Internet session. This criterion will also attempt to measure the rate of overblocking, or the blocking by mistake of harmless websites.

Efficiency – This criterion refers to the speed with which the particular technology functions. Does it considerably slow down the browsing? In addition, it involves the analysis of the requirements of the particular system, in particular, the consumption of RAM memory.

Ease of integration – This refers to whether the system counts with a mechanism to be integrated with other applications. This mechanism could be a library, an API (Application Program Interface) or a finished product.

Degree of development of the technology – This criterion refers to whether or not there is a finished product based on the technology. In addition, it will also attempt to measure the stability of the product, i.e., how well does the system work without crashing.

Languages supported - The project is a European project, therefore it is important to learn whether the system deals with harmful contents in different languages. This evaluation makes sense only for those techniques that use some kind of text analysis

4.3.2 ENVIRONMENT

In order to ensure the neutrality of the study it is crucial to have each technology tested in the same technical environment. Therefore, we suggest that the evaluation of each of the different tools should be carried out in Pentium 400 MHz machine.

4.4 RESULTS OF THE TESTING

The next sections present the result of the testing done with the different tools, products and prototypes according to the criteria explained in the previous section. Regarding Filterix, what the Demokritos Institute kindly did was to give the NETPROTECT consortium access to their internal proxy server version since they are in the process of commercialising their filtering solution and giving away an executable, even for testing purposes would be impossible for them. This solution allowed us to test overblocking and underblocking and to be able to have a fair idea on Filterix's behaviour with respect to content. However, we were not able to judge, for example, service speed (due to Internet latency) and ease of installation on various platforms. In addition, it is important to stress that OPTENET is part of the Netprotect consortium.

4.4.1 EFFECTIVENESS

The rate of effectiveness and the overblocking were computed taking into account the results of the analysis of the tools, products and prototypes with checkprotect. The rate of effectiveness is a comparison between the number of harmful URLs used in the analysis and the number of URLs the tool, product or prototype actually blocked. The rate of overblocking was obtained by comparing the total of safe URLs that did not need to be blocked with the number of URLs in this list that were actually blocked by the tool,product or prototype. A detailed explanation for each tool is provided in the sections below.

Tools, products and prototypes that use text-based content analysis techniques:

	RATE OF EFFECTIVENESS	**OVERBLOCKING**
OPTENET	60,1%	6,2%
THE BAIR	60,5%	18,12%
PURESIGHT	63,5%	9,66%

Tools, products and prototypes that use image analysis techniques:

	RATE OF EFFECTIVENESS	**OVERBLOCKING**
FILTERIX	45,4%	0,59%

Table 1: Effectiveness

This rate was measured with the same list of URLs used in the testing of the filtering products.

4.4.2 EFFICIENCY

	OPTENET	THE BAIR	PURESIGHT
How much does it slow down the browsing? (Rate on a scale of 0 to 3)	0	1,5	1
What is the system's consumption of RAM memory? (Answer in Mbytes)	2,7 Mb Swap 15,7 Mb	4,180 Mb	4Mb

Table 2: Efficiency

4.4.3 EASE OF INTEGRATION

	LIBRARY	API	FINISHED PRODUCT
OPTENET			✔
THE BAIR			✔
PURESIGHT			✔

Table 3: Ease of integration

4.4.4 DEGREE OF DEVELOPMENT OF THE TECHNOLOGY

	OPTENET	THE BAIR	PURESIGHT
Does the system crash? (Answers should be: never, sporadically, several times during a work day)	Never	Sporadically	Sporadically

Table 4: Degree of development

4.4.5 LANGUAGES SUPPORTED

	English	Spanish	French	Portuguese	Italian	German	Autolearrn of new languages
Optenet	✔	✔	✔	✔	✔	✔	
The Bair	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Puresight	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 5: Languages

5 TOOLS, PRODUCTS AND PROTOTYPES THAT USE TEXT-BASED CONTENT ANALYSIS TECHNIQUES

The consortium has identified the following tools, products and/or prototypes that use text-based content analysis techniques. A brief description of the different tools, products and prototypes is provided below. It is important to emphasise that the information on the effectiveness rate as well as all the other information in this section is taken directly from the different companies' web sites.

5.1 OPTENET:

Web page of product: www.optenet.com

Brief description of product: OPTENET is based on two components: the predefined lists and the analysis engine. The analysis engine analyses the contents of the web page that is being consulted and decides whether it has to be blocked or not. Just before the web page is shown in the browser the analysis engine examines it performing a semantic analysis based on contextual topic recognition. In this way, it decides whether the page will be shown or not. Currently, the analysis engine can analyze pages in several languages: Spanish, English, French, Italian, German and Portuguese.

Operating Systems supported: Windows 95, Windows 98, Windows Millenium, Windows NT Workstation, Linux, Solaris

Where to find demo: www.optenet.com

For the evaluation carried out in this report OPTENET prepared a special version of its tool. As it was explained earlier in the report OPTENET's tool is based on two components: the predefined lists and the analysis engine. For the purpose of the testing of the different techniques used in filtering carried out in this report, OPTENET prepared a version of its tool which excluded the predefined lists. This special version of the tool was composed only of the analysis engine used by the tool. OPTENET is part of the NetProtect consortium.

The test of the efficiency of OPTENET has given the following results:

- 1622 of the 2890 URLs to be blocked were filtered (56,1%)
- 88 of the 1659 URLs not to be blocked were filtered (5,3%)

In English OPTENET obtained an efficiency rate of 59,47% (386 of the 644 English URLs to be blocked were blocked).

In the French list of URLs its results were: 311 out of the 566 URLs to be blocked were actually blocked (54,94%). Regarding the German URLs, it blocked 411 of the 662 harmful URLs (62,08%).For Spanish it blocked 423 harmful URLs out of 665 (63,60%).

Like most of the other tools, systems and prototypes tested it obtained its worst results in the blocking of the Greek pages, only 26,62%.

5.2 RULESPACE

Web page of product: www.rulespace.com

Brief description of product: The Contexion Embedded Analysis Toolkit (EATK) is a set of software libraries with application programming interfaces that provides real-time Web content recognition.

The EATK uses patent-pending techniques to extract the unique features (patterns of letters, words and phrases) of a given web page. Then these patterns are compared to a pre-trained category model, and if the features of the web page match those of the category model, the page is determined to "belong" to that category.

Operating Systems supported: Information not available on web site

Where to find demo: Demo versions of the product can be requested on the following URL: www.rulespace.com/forms/demo/

Rulespace Inc. provided the consortium with a paper entitled *"Contexion Services – Technical Overview"* from September of 2000, explaining their technology. Highlights are provided below.

5.2.1 WHAT IS CONTEXION SERVICES?

Contexion Services is an embeddable technology and Internet infrastructure service that enables applications to automatically recognise the subject matter of a web page or web site. Its technology emulates the human process of inferring the meaning or intent of digital information by considering both the features of the content and the context in which the features exist. According to its producers, this permits the creation of applications that can detect, filter, profile, organise and discover web content.

Contexion Services utilises an ontology expandable to thousands of pre-defined subject categories that reflect the distribution of web information. Each category is represented as a unique node within a multi-category classification system constructed using neural network technology. In a real-time application, buffers of unknown digital information are submitted to the classifiers, which in turn associate the content to one or more of the categories within the ontology.

Contexion Services is accessible through two different but complimentary software development toolkits. The Embedded Analysis Toolkit enables categorisation of real-time, dynamic media stream directly within an application. The Online Lookup Toolkit provides applications access to an online service that processes category ratings of public web sites and their subdirectories.

5.2.2 HOW DOES CONTEXION TECHNOLOGY WORK?

The basic unit of Contexion technology is the *category model*: a category detector used to identify an unknown buffer of text (i.e., a web page) as a member of a specific category within the WorldView ontology. For example, three separate category models represent the category Leisure/Sports/Football: one for Leisure, one for Sports and one for Football.

A category model is the output of a patent-pending process using neural networks to extract the unique features that exist within the content of a given category. Contexion is based on pattern recognition – not semantic- or linguistic-based techniques – enabling Contexion Services to recognise web content written in any language given sufficient content examples to adequately train a category model.

5.2.3 CATEGORY MODEL TRAINING

At a very high level, a category model is built by presenting the following to a neural network: samples of the *target* category and samples of the *anti-target* content that represents the Web at-large. Through the *training process*, the neural network learns to distinguish between the target and anti-target content – learning what is and what is not the given category. Once a sufficient level of accuracy is attained, the resulting set of features becomes the model for the category within Contexion Services.

5.2.4 CATEGORY MODELS IN ACTION

Contexion Services uses the category model as the basis of reference when recognising unknown web content. The features of an unknown web page are extracted in real-time. Dot-product mathematics are then applied to the feature vector weights of the unknown content to determine their significance with respect to those of the category model. If a significant degree of similarity exists, the web page is identified as belonging to the category represented by the category model.

Whether recognising the content of an individual page or analysing all the pages contained within a web site or its subdirectories, the category model serves as a vital foundation to Contexion Services.

5.2.5 CONTEXION SERVICES WORLDVIEW

The Contexion Services WorldView is an ontology of subject matter used to define and organize the categories available within the Contexion Services. This categorisation scheme reflects the stores of information contained within the World Wide Web in the same way that the Dewey Decimal System does for a conventional library. Unlike traditional Library Science schemes though, the World View is specifically oriented to web information. Its top-level categories include prevalent web content such as Computers, Entertainment, Health, Leisure and Shopping, as well as the more conventional categories such as Arts and Sciences.

5.3 PURESIGHT (ICOGNITO)

Web page of product: www.icognito.com

Brief description of product: PureSight is based on Artificial Content Recognition (ACR) technology. ACR uses artificial intelligence algorithms that mimic the human ability to understand material and categorize it according to its content. iCognito's ACR technology is multilingual, and can learn any language or category in a short period of time.

Operating Systems supported: Windows 95, 98, ME, 2000 and NT.

Where to find demo: A thirty day trial version of the product can be requested on the web page of the product: www.icognito.com

The test of the efficiency for PureSight (iCognito) has given the following results:

- 1874 oft the 2951 URLs to be blocked were filtered (63,5%)
- 164 of the 1697 URLs not to be blocked were filtered (9,66%)

The tool is very easy to install, it allows the user to include lists of keywords in order to adapt it to each users' own needs.

It slows down moderately the speed of navigation in a Pentium 400, in inferior machines this is more noticeable.

In the tests conducted, it has shown good stability in a machine that carries out normal Office tasks with scarce navigation. However, it has produced the crashing of the CheckProtect application many times. Therefore, it seems the tool produces instability on heavy Internet usage situations.

Concerning languages, it has answered well to all of the languages except Greek. However, its web site does not provide any information on the languages supported by the system. It mentions that the system can rapidly "learn" other languages, nevertheless there is nothing in the product that indicates this.

Puresight obtained its best results in German 536 out of 695 German URLs to be blocked were blocked (77,12%).

With the English list it obtained an efficiency rate of 69,89% (448 of the 641 English URLs to be blocked were blocked).

In the French list of URLs its results were: 334 out of the 560 French URLs to be blocked were actually blocked (59,64%). For Spanish it blocked 414 harmful URLs out of 658 (62,91%).

Like most of the other tools, systems and prototypes tested it obtained very poor results in the blocking of the Greek pages, only 142 out of 397 harmful pages were blocked (35,76%).

5.4 SAIL LAB'S AUTOMATIC TEXT CLASSIFICATION SYSTEM

Web page of product: www.sail-labs.de

Brief description of product: Sail Labs has been working on automatic text classification for several years. In [2] an experimental comparison of various methods and approaches is presented. Sail Labs system for automated text classification has recently been finished. It is the essence of several previous prototypical systems and uses the newest techniques from information theory and neural networks: Conditional Mutual Information, Support Vector Machines, Non-Linear Classification. The system is able to identify higher-order relationships between words (and phrases) in order to effectively disambiguate word occurrences without the use of semantics. It is completely language-independent. An application for classifying news agency articles (hierarchical topic structure consisting of 120 topics) has been realised for six languages. Sail Lab´s text classification system is very fast. In the news classification application 200 MB of text can be analyzed (classified) in one hour.

Sail Labs i part to the Netprotect consortium.

Operating Systems supported: Windows NT, 98, 2000, Linux

Where to find demo: currently there is no demo version available.

5.4.1 THE TECHNOLOGY

Sail Labs has been working on automatic text classification for several years and they are part of the NetProtect consortium. In [2]] an experimental comparison of various methods and approaches is presented. Sail Labs system for automated text classification has recently been finished. It is the essence of several previous prototypical systems (described in [2]) and uses the latest techniques from information theory and neural networks.

Sail Lab´s system for automatic text classification comprises three main steps (see also Section 2.3.3.): *feature construction, feature selection, and the actual classification step.*

5.4.2 FEATURE CONSTRUCTION

Sail Lab´s text classifier uses a brute force approach almost without any linguistic analysis. As features, individual *words* (a very simple language-independent tokenizer is used) and/or *letter n-grams* (letter sequences) up to a fixed length are used. The maximum n-gram length can be configured and lies typically between 6 and 10. For the generation of letter n-gram features word boundaries are ignored, so *inter-word* letter n-grams which represent multi-words (e.g. the multiword 'New York' becomes an n-gram feature of length 8) are also generated. It is this simple feature construction step that makes the system completely language-independent.

A preliminary inspection of pornographic web pages has shown that such pages often contain non-lexical expressions (special expressions not known in standard lexicons) and that important hints about the pornographic contents can be found in substrings of links or names of figures. This causes problems for systems that build on a linguistic analysis of text. However, the non-linguistic brute force approach of Sail Lab´s system can cope with such text-properties very well.

5.4.3 FEATURE SELECTION

The letter n-gram approach for feature construction can produce several millions of sometimes highly correlated features for a training corpus. Therefore a highly sophisticated and efficient feature selection method is needed. Sail Lab´s text classifier uses a topic-specific feature selection mechanism based on the *conditional mutual information measure*. With mutual information it can select those features that carry most information for identifying a topic. However, conditional mutual information not only measures the information about a topic gained by one feature alone. It also measures how much a feature contributes together with up to two other features. Thus the system is able to recognize that e.g. the feature “breast” together with “boneless” or “chicken” indicates the topic cooking while together with “feeding” or “pump” it indicates a medical topic.

5.4.4 CLASSIFICATION

While the feature selection step selects features that carry much information for identifying a topic, the actual classification step builds a model for a topic. Roughly speaking, such a model determines how many relevant features and in which combination these features have to occur in a text in order to associate the topic with the text. The model may also include rules stating that specific features or combinations of them must not occur in texts that are associated with the topic. Sail Lab´s text classifier uses neural networks that is *single- or multilayer perceptrons* (non-linear classification) for the actual classification step. However, the system does not use the standard backpropagation gradient algorithm from the mid 80s. It achieves a much faster convergence by using a conjugate gradient algorithm. Furthermore, Sail Lab´s text classifier avoids overfitting very successfully by applying a sophisticated model complexity control method (very similar to what support vector machines do) that determines the appropriate model complexity for a topic before the actual training starts.

5.4.5 APPLICATIONS AND RESULTS

An application for classifying news agency articles (hierarchical topic structure consisting of 120 topics) has been realised for six languages. The average results lie around 75% (average of recall and precision) varying considerably from topic to topic. In general one can say that the effectiveness depends heavily on the quality and number of training data. Usually several hundred example documents are needed per topic. In order to relate the results to the pornographic filtering application we can state the following: If the task were e.g. to filter text on economic subjects the system currently has an effectiveness of 89% with an overblocking of 8% on Reuters news stories.

5.4.6 KEY FEATURES

The key features of Sail Lab´s text classification system can be summarised as follows:

- Fully automatic text classification for filtering and routing applications: users have to provide example text for the topics they are interested in; no formal definition of topics necessary; the system extracts (learns) the relevant concepts automatically and is then able to identify the topics in new documents.
- Latest technology from the fields of Information Theory, Statistical Learning Theory, Neural Networks, and Patter Recognition.
- Language-Independence: since the system currently does not apply linguistic knowledge, it can be used for any language without adaptation. However, training has to be done for each language separately.
- Format-Independence: supports Otext and consequently all formats for which Otext-Transformation exists (currently only HTML**).**
- Topic Organization**:** the system is able to handle topic hierarchies / ontologies; more than one topic may be assigned to one document.
- Classification of new documents is very fast: about 200 MB of text can be classified per hour. Consequently, if the system is integrated as filter into a browser, the slow-down can be neglected in comparison to the time needed to download the web pages.

- The system computes relevance-rankings instead of boolean classifications: a document may get a 95% rating for shopping, 80% for cars and 12% for sport; this means that the user has control over the recall / precision trade-off.

- The system is available as library with an API. It runs on Windows NT, 98, 2000, and Linux.

5.4.7 APPLICATIONS AND RESULTS WITH SAIL LABS CLASSIFIER

An application for classifying news agency articles (hierarchical topic structure consisting of 120 topics) has been developed for six languages. The average results lie around 75% (average of recall and precision) varying considerably from topic to topic. In general one can say that the effectiveness depends heavily on the quality and number of training data. Usually several hundred example documents are needed per topic.

The first experiments on the English URL collection have already been carried out. For the training 369 pornographic URLs and 110 non-pornographic URLs (that had been collected by the consortium), plus some 5000 Reuters news agency articles were used (the classifier needs examples and counter-examples for training). Training and testing was done with five-fold cross validation. This means that the whole data set was partitioned into five parts of equal size, four parts were used for training and one for testing and this was done for all possible 5 perturbations. The final results that are given are averaged over the five folds: 97% effectiveness, 0.3% overblocking on all counterexamples, 13% overblocking (if only measured on the specially collected non-pornographic URLs).

6 TOOLS, PRODUCTS AND PROTOTYPES THAT USE IMAGE ANALYSIS TECHNIQUES

The consortium identified the following tools, products and prototypes that use image analysis techniques. A brief description of each of them is provided below. It is important to emphasise that the information on the effectiveness rate as well as all the other information in this section is taken directly from the different companies' web sites.

6.1 EYEGUARD

Web page of product: www.eye-t.com

Brief description of product: Using image analysis, Eyeguard checks the images being displayed for excessive skin tones, thereby protecting the user from pornographic images. Once installed, explicit images displayed on the screen from any source will automatically be blocked.

Operating Systems supported: Windows 95, Windows 98, Windows NT.

Where to find demo: There is no demo available for download on the web site, however those interested in evaluating Eyeguard, should contact their Sales Team on +44 (0) 1274 530748.

6.1.1 THE COMPANY

EYE-t Technology Limited was founded in early 1996 with a directive to develop innovative software solutions for the IBM and compatible PCs. The company has a product range to suit the business, education, and government market.

6.1.2 THE TECHNOLOGY

Using image analysis, Eyeguard checks the images being displayed for excessive skin tones, thereby protecting the user from pornographic images. According to the company once the program is installed, explicit images displayed on the screen from any source will automatically be blocked - immediately. The PC will then remain unusable until the supervisor unlocks it. Alternatively, Eyeguard can act as a watchdog, detecting unacceptable images silently, leaving the PC user unaware. In both cases the relevant images are stored in a central log, so that the teacher or IT manager can examine the audit trail of each users activities.

6.1.3 KEY FEATURES

According to EYE-t these are the major features of their product (taken from their web site):

- Detects images from any source:- CD-ROM, Floppy Disk, Hard Disk, Internet, Network, E-mail, etc.

- Works with photographs and movies
- Supports all video resolutions and screen depths of 256 colours or more
- Extremely fast detection engine
- Highly visible Active Mode lockout screen message can be customised
- All trigger events logged and thumbnails saved for viewing via the eyeguard control panel
- Detection engine can be easily enabled or disabled (even for a set length of time) using the eyegard control panel or the eyeguard network control panel
- User configurable detection parameters.

Password protected control panel as standard for added security.

6.2 FILTER SOFTWARE FIRST 4 INTERNET

Web page of Product: http://www.first4internet.co.uk/about.htm

Brief description of product: The filter analyses the content of an image and determines if it has attributes of pornographic nature. The analysis incorporates artificial intelligence while determining attributes, orientation and content of each image in order to assess whether or not an image is pornographic. F4i has constructed a library of analytical capabilities, which provides the filter with its ability to process digital images. It can also be given profiles that represent/recognise other images such as company logos, trademarks or offensive icons. The filter operates without having to rely on large server based databases, the software footprint is less than 5Mb. The application is cross platform enables with the ability to reside either Server side on an ISP or Intranet, or be supplied as a standalone PC product.

Operating systems supported: Information not available in web site.

Where to find demo: A trial version can be obtained by requesting it at http://www.filter.first4internet.co.uk/

6.2.1 THE COMPANY

First 4 Internet Ltd (F4i) appeared as a software development business providing software solutions for businesses targeting the mass market. They are specialised in Image Content Filtering and Image Composition Analysis (ICA) software, however they also offer other security solutions. Recently, F4i is cooperating with Surf Control putting their Image Content Filtering at the disposal of the Surf Control filtering solution.

6.2.2 THE TECHNOLOGY

The filter software combines image analysis and text content analysis in order to prevent offensive or adult images and pornographic text content from reaching the computer screen. This comprehensive filter system carries out a number of detailed permutations at high speed in order to assess the suitability of each image for viewing. Through the combination of 22,000 algorithms (as First 4 Internet explains on their web page - that means 22,000 analysis steps), the image content analysis software should distinguish between images of an artistic nude and those that are pornographic, with a high degree of accuracy. The ICA technology uses multiple methods of scanning files for pornography.

6.2.3 APPLICATION

The application is cross platform enabled with the ability to reside either Server side on an ISP or Intranet, or be supplied as a standalone PC product. F4i will license their image categorisation and learning capability software package to SurfControl for the purposes of development and distribution by SurfControl of an "image engine" software package enabling customers to automate the categorisation of images contained in computer data files.

Initially, the ICA technology has been developed to cope with attachments of emails, so that users can apply this filtering system in protecting mailing programmes against receiving emails with pornographic content.

6.2.4 TESTING

The technology has been tested by the company TesCom (http://www.tescom.co.uk), an independent software testing company, which concluded that the ICA software prevents 95 percent of pornographic images from reaching the computer screen. TesCom also found that the software is able to scan more than 20 images per second, using a test sample of 3,000 images on a server. The firm says that similar tests reveal that other technologies can only achieve an accuracy of 67 percent. Unfortunately, TesCom does not indicate the rate of overblocking. The whole test report of TesCom was not available.

6.3 LOOKTHATUP.COM

Web page of Product: http://www.lookthatup.com

Brief description of product: LookThatUp Image-Filter is a server-side software that is able to understand and subsequently filter out offensive images over Internet. This software automatically sees and analyses the content of images and compare it with is database of images. After the comparison between the tested image and the database images LookThatUp give a pornscore of the tested image according to its degree of pornography.

Operating systems supported: Windows 95, Windows 98, and Windows 2000.

Where to find demo: There is no demo available for download on the web site, however those interested in evaluating LookThatUp, should contact their Sales Team on sales@lookthatup.com.

6.3.1 ABOUT THE COMPANY

LookThatUp was founded in 1999 by a team of scientists from the Image and Multimedia Indexing Group at INRIA, the French National Institute for Research in Computer Science and Automatics. This group developed a highly sophisticated system of algorithms based on image techniques to analyse, indexes and retrieves images. One of their most popular system is SurfImage, a "user-friendly, generic and flexible content-based image retrieval system". This system based on a query-by-example approach allows the user to retrieve pictures which are visually similar to a reference picture. LookThatUp has further developed this technology and now propose tools for searching or filtering images.

6.3.2 ABOUT IMAGE FILTER

Image Filter is a server-side software which works as a proxy being able to filter out harmful images over the Internet. Image Filter XWatch deals with the filtering of pornographic images.

Image Filter XWatch analyses images submitted by a process of segmentation and extracts a signature of the image composed of different visual descriptors (colour, texture, shape and a composite descriptor). This signature is then compared with other signatures previously computed on the images contained in a large database of the LookThatUp server.

At the end of this process (which takes 0.6 second), the system outputs a score for the submitted image between 0 and 100, a score of 0 correspond to a totally harmless image and a score of 100 corresponds to a highly pornographic content. The end-user of the system can define himself/herself the level of acceptance (e.g. for a level of acceptance of 50, every image with a score higher than 50 would be blocked).

6.3.3 EXPERIMENTS AND RESULTS

In order to assess the efficiency of the approach, LookThatUp granted us with a free copy of a client version of Image Filter XWatch Version 1.1.1 beta. The only purpose of this client version is for testing the efficiency of the server-side software. To our knowledge, there is no standalone copy for direct use by parents/teachers as a filtering tool. Image Filter presents itself as a server-side software working through the Internet and easy to integrate in other software. As such, it is perfectly fit with the objectives of the NetProtect project.

The installation of the testing client was very easy and very fast (2 minutes). There was no need for special configuration and the testing client was ready and easy to use. The user is invited to select a directory, which contains the pictures to test. Note that the Image Filter XWatch only works with pictures and not with synthetic images, banners, animated GIFs or buttons.

Several tests were carried out in order to get an idea of the efficiency of LookThatUp image filtering technique.

A first the test includes two series of images: one series of 100 harmful images and one series of 100 harmless images. Each image weighted around 40 Kb.

First of all it appeared that filtering an image is very fast. This was one of our major worries since the system works as server-side software. Each picture therefore has to be sent over the Internet in order to be analysed and a score is then returned. It appeared that the test of each series of 100 pictures took around 10 minutes in total, which indicates an average time of 6 seconds. Testing the system on a RTC connection with a 28.8 KBPS modem showed average times of 15 seconds. However, on smaller images (10 Kb), the results were in near real-time without any impact on the classification. On the contrary, on larger images, the response time were higher and up to 90 seconds.

Our test showed that black and white pictures were detected. Some of these pictures were analysed and filtered and others not.

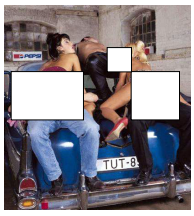
The first series of 100 objectionable images gave the following results. 27 out of the 100 pictures were not analysed. LookThatUp explained to us that changes were made on the server and that a new copy of the testing client was needed in order to take these changes into account. This however does not change the quality of the analysis made on the 73 other images. The following table shows the score returned by Image Filter on 73 pornographic pictures. A score of 100 should indicate an image with "highly pornographic content" while a score of 0 should indicate a "totally harmless image".

Score	**≤ 30**	**30-40**	**40-50**	**50-60**	**60-70**	**≥ 70**
Percentage of images	3%	1%	8%	4%	9%	75%

Table 6: Scores on objectionable pictures

The analysis of these results shows the effectiveness of filtering harmful images: 75% of these pornographic images have a score higher than 70 and 90% of these pornographic images have a score higher than 50. The following figure gives an example of a picture which scored less than 30 and could therefore be interpreted as harmless though they should be classified as pornography.

Porn score: 12



Porn score: 21

Porn score: 30

Figure 2: Objectionable pictures with low porn scores

The second series of 100 images contained various images including nudity (6 pictures), lingerie (50 pictures) but also pictures of animals (7 pictures), people (11 pictures). One of the objective of the test was to test the ability of the system to make the difference between pornography and nudity, between pornography and lingerie. Lingerie pictures can typically be found on electronic commerce websites which certainly does not have to be filtered.

Score	≤ 30	30-40	40-50	50-60	60-70	≥ 70
Percentage of images	38%	16%	7%	4%	11%	24%

Table 7: Scores on potentially harmless pictures

The results for these potentially harmless images was less positive: 40% of images have a score higher than 50 and could therefore considered as objectionable.

Some pictures got a high score though they were clearly harmless: a picture of a doe with its fawn scored 76 and a picture of waterfall scored 52.


Porn score: 76


Porn score: 78

Porn score: 85

Porn score: 100

Figure 3: Potential harmless pictures with arguing scores

The results on the lingerie pictures were also a bit difficult to interpret: some scored high and others scored low.

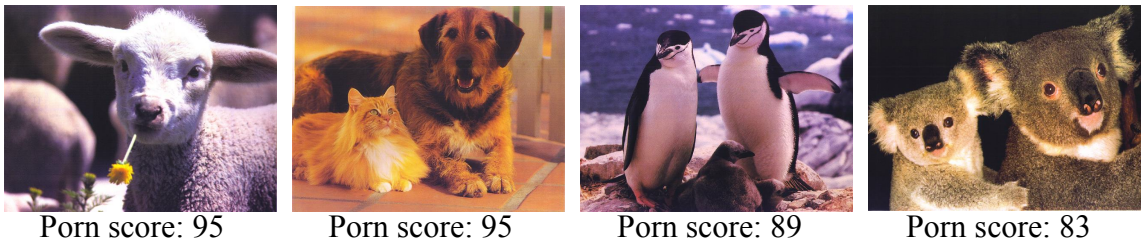
Our first conclusion was therefore a tendency of the system to overblock pictures.

It was therefore decided to conduct another test on a series of 100 pictures. This series only contained pictures of animals, scenic views (e.g. Grand Canyon, Yosemite National Park, Niagara Falls, etc.), famous places (Capitol, London Parliament, etc.). The commonality between all these pictures is that there is no people photographed and no one could argue about the fact that these pictures are harmless.

Score	≤ 30	30-40	40-50	50-60	60-70	≥ 70
Percentage of images	80%	6%	1%	4%	0%	8%

Table 8: Scores on harmless pictures

The results of this third series were a lot more encouraging. It was remarkable how the system globally identified that this series of pictures did not contain pornography. The distribution of the scores is quite the opposite of the distribution made on the first series of pictures (see Table 6). One will note that they are still harmless pictures, which are ranked with very high scores though they do not exhibit pornography.



Porn score: 95 Porn score: 95 Porn score: 89 Porn score: 83

Figure 4: Harmless pictures with high porn scores

Another round of test was run with a second version of LookThatUp Image Filter Xwatch v.1.1.2beta. This second round of test was conducted on the same pictures as the first round of test (100 harmful pictures and 100 harmless pictures).

The results of this second round of tests are similar to the first except that in the second round, all pictures were analysed. The scores of the pictures do not change from the first version on Xwatch. The scoring process is constant for the same pictures.

Score	≤ 30	30-40	40-50	50-60	60-70	≥ 70
Percentage of images	1%	4%	2%	5%	1%	83%

Table 9: Scores on harmful pictures (new version of Xwatch)

Score	≤ 30	30-40	40-50	50-60	60-70	≥ 70
Percentage of images	79%	6%	1%	5%	2%	7%

Table 10: Scores on harmless pictures (new version of Xwatch

6.3.4 SUMMARY OF FINDINGS

The tests that we have conducted show that LookThatUp filtering solution can really bring an added-value to an integrated solution as the one we intend to develop in the context of the NetProtect project.

In order to get better filtering results, one might set up 2 limits. The first limit could be set at 30. All images returned with a score of 30 of lower could be considered as harmless with a very limited risk of error. The other limit could be set at 70. All images returned with a score of 70 or higher could be classified as porn pictures. If we consider this approach on the series of images we considered, we get the results presented in the next table.

Score	Errors	Not classified	Correctly filtered
Porn pictures	3 %	22 %	75 %
Potentially harmless pictures	24 %	38 %	38 %
Harmless pictures	8 %	11 %	80 %

Table 11: Summary

In addition to this recommendation, it is worth taking into account the fact that LookThatUp solution analyses image one by one. However, it is often the case that web pages display several pictures. Porn web sites will often display galleries, which contain several porn pictures. The classification of the page will therefore require the analysis of all the images contained in the page. Since this solution performs well on porn pictures and harmless pictures, it will be easy to combine the scores of all images contained in the web page in order to calculate an overall score.

Figure 5: Harmful website

A short experiment conducted on a set of thumbnails all coming from the same web page (see Figure 4) with pornography material was very positive. All 7 thumbnails scored higher than 50 and 3 scored higher than 70 in less than 1 minute.

The only limitation to this approach can be the time that will be needed to get all the scores.

6.4 EASYGLIDER

Web page of product: http://easyglider.com

Brief description of product: EasyGlider is a multimedia navigation software product, based on highly innovative and very fast pattern matching algorithms, implementing a new approach for information retrieval from multimedia documents. EasyGlider builds links among multimedia documents. These links enable users to navigate intuitively and quickly in multimedia database.

Operating Systems supported: Linux Red Hat 6.2, Windows NT4, and Windows 2000.

Where to find demo: http://www.easyglider.com/demo.htm

The company was found in June 2000.

EasyGlider provides a new generation of multimedia navigation engine.

EasyGlider is a multimedia navigation software product, based on highly innovative and very fast pattern matching algorithms, implementing a new approach for information retrieval from multimedia documents.

The system is built on a set of light components, which can provide indexing, and navigation services. The system is based on a graphical analysis of images and a semantic analysis of text. EasyGlider analyses the visual components of images, by evaluating similarities based on colour, texture and shape, to index pictures. For textual information, EasyGlider analyses concepts and creates indices from semantic similarities. Thus EasyGlider builds links among multimedia documents. These links enable users to navigate intuitively and quickly in multimedia database.

In fact, EasyGlider indexes each document and built links between similar documents. In this way, user can navigate in the database of multimedia documents by clicking on images like hypertext links from each document to its neighbours.

6.5 WIPE WAVELET IMAGE PORNOGRAPHY ELIMINATION

Web page of Product: http://www-db.stanford.edu/IMAGE

Brief description of product: This system is capable of classifying a website as objectionable or benign based on image content. The system compares tested images with a database of classified images.

Operating systems supported: Information not available on web site.

Where to find demo: an online demo is available at http://wang.ist.psu.edu/cgi-bin/zwang/wipe2_show.cgi

6.5.1 ABOUT THE AUTHOR

James Z. Wang has assisted in teaching mathematics at both Stanford and Minnesota University. Additionally, he has assisted in teaching biomedical informatics at Stanford, covering topics such as computational molecular biology, DNA sequence analysis, and medical imaging. Wang is an expert in visual database search and retrieval formerly with Biomedical Informatics Group and the Computer Science Database Group at Stanford, he undertakes work that makes possible the retrieval of specific images from databanks of photos in the first time for medical uses. After, with Gio Wiederhold, they created the IBCOW system (Image-based Classification of Objectionable Websites), a system capable of classifying a website as objectionable or benign based on image content. This system uses the WIPE technology, a technology developed by these two authors.

6.5.2 ABOUT WIPE TECHNOLOGY

The algorithm used in this technology is a combination of Daubechies wavelets and colour histograms to provide a semantically meaningful feature. This feature is compared with pre-marked images as objectionable or benign in an image database (containing more than 200000 images). The system is very fast and practical for real time applications, processing queries at the speed of less than 10 seconds each, including the time to compute the feature vector for the query.

For IBCOW application (Image-based Classification of Objectionable Websites), if more than a certain number of images sampled from a site are found to be objectionable, then the site is considered to be objectionable.

An on-line demo of the WIPE technology can be tested on the web: http://wang.ist.psu.edu/cgi-bin/zwang/wipe2_show.cgi

Several studies were made by Wang and al. on the WIPE technology: in the first study [4], the system identifies objectionable pictures with 97.5% hits over a test set of 437 images found from objectionable news groups, it wrongly classified 18.4% of a set of 10809 benign images obtained from various sources. In the second study [5], the system shows 96% of sensitivity over a test set of 1076 digital photographs found on objectionable news groups, the system wrongly classified 9% of a set of 10809 benign photographs obtained from various sources.

7 TOOLS, PRODUCTS AND PROTOTYPES THAT USE BOTH IMAGE AND TEXT BASED CONTENT ANALYSIS

The consortium has identified some tools, products and prototypes that use both image analysis techniques and text based content techniques.

7.1 THE BAIR FILTERING SYSTEM

Web page of product: www.thebair.com

Brief description of product: The Bair filter was developed and trained to block pornographic images and text from the Internet. The Bair's image recognition filter has been taught to recognise sexually explicit graphics regardless of the text on the Web page and it is used in conjunction with the Bair's text filter.

Operating Systems supported: Windows 95, Windows 98, Windows 2000, and Windows NT.

Where to find demo: www.thebair.com

The test of the efficiency of The Bair filtering system has given the following results:

- 1820 of the 3004 URLs to be blocked were filtered (60,5%)
- 306 of the 1689 URLs not to be blocked were filtered (18,12%)

The tool is very easily installed. It can be downloaded from its web site. It gives the user the option of filtering images or not.

It is not a local application, it is installed in a remote proxy. Consequently, it does not consume the system's resources. On the other hand, the navigation speed is noticeably reduced.

Regarding languages, The Bair Filtering System obtained its best results in English with an efficiency rate of 77,58% for this language (488 of the 629 English URLs to be blocked were blocked).

In the French list of URLs its results were: 369 out of the 564 URLs to be blocked were actually blocked (65,42%). Regarding the German URLs, it blocked 467 of the 766 harmful URLs (60,96%). For Spanish it blocked 249 harmful URLs out of 673 (51,85%).

Like most of the other tools, systems and prototypes tested it obtained very poor results in the blocking of the Greek pages, only 39,51%.

7.2 FILTERIX

Web page of product:

http://www.iit.demokritos.gr/skel/en/Projects/FILTERIX_Description.htm

Brief description of product: Filterix is a prototype software tool for the blocking of obscene content accessible through browsers on the World Wide Web. The current implementation demonstrates the technology results by offering a Web filtering proxy service. Filterix combines technologies such as machine learning, language learning and image processing, in order to achieve real-time analysis and further blocking of pornographic content on the Internet. In the current implementation the analysis process is based on real-time filtering on the content of the URL requested by the client browser. The type of information, fused under a probabilistic framework, draws from the text, images and metadata of each page so as to reach an overall probability that the page is pornographic or not. If it is identified as such, the page is blocked and an explanatory message is sent to the user.

Operating Systems supported: All operating systems since all the code is written in Java.

Where to find demo: Contact Demokritos National Centre for Scientific Research, http://www.itt.demokritos.gr/

Since the National Center for Scientific Research (Demokritos Institute) is in the process of commercialising their filtering technology, giving away an executable, even for testing purposes, understandably presented a problem to them. However, they provided the NETPROTECT project access to their internal proxy server version, which uses basically the same technology. This gave the consortium in NETPROTECT the opportunity to test overblocking and underblocking, and also to have a fair idea on the prototype's behaviour with respect to content. However, we were not being able to test, for example, service speed (due to Internet latency) and ease of installation on various platforms.

The test of the efficiency for Filterix has given the following results:

- 1339 of the 2951 URLs to be blocked were filtered (45,4%)
- 10 of the 1687 URLs not to be blocked were filtered (0,59%)

As it can be seen, the rate of overblocking was surprisingly low for Filterix. It got its best results for blocking in the English language with an efficiency rate of 61,38% (399 of the 650 English URLs to be blocked were blocked).

In the French list of URLs its results were: 236 out of the 584 URLs to be blocked were actually blocked (40,41%). Regarding the German URLs, it blocked 363 of the 751 harmful URLs (48,33%). For Spanish its results were similar to that of French 284 harmful URLs were blocked out of 683 (41,58%).

Like most of the other tools, systems and prototypes tested it obtained very poor results in the blocking of the Greek pages, only 20,14%.

7.3 WEB WASHER EE

Web page of product: www.webwasher.com and www.cobion.de.

Brief description of product: The WebWasher Enterprise Edition (EE) of the webwasher.com AG (part of the Siemens AG) is a server-based product that intends to optimise Internet use in the workplace, libraries and schools by combining Internet Content Filtering and Internet Access Management. The Web Washer EE uses image recognition methods (Cobion technology) as well as text recognition methods. The results of these two types of filtering techniques are collected in a big database and reviewed by human interaction.

Operating Systems supported: Windows NT Server, Windows 2000 Server, Linux and Solaris

Where to find demo: http://www.webwasher.com/en/products/wwash/download.htm

Combining Internet Content Filtering and Internet Access Management WebWasher filters by object size, uses the filtering of URLs and strings, pop-ups, pictures, applets, plug-ins, advertising banners, scripts and animations (GIF, Shockwave, Flash), deals with negative and positive selection of filter criteria, filters cookies and Web bugs; i.e. WebWasher does not have only blocking functions. It respects also security criteria. A Media Type Filter blocks risky file types before they get into the corporate network. All current protocols like html, email, newsgroup, chat can be blocked.

The administrator can control the access and the filtering functions by remote administration via a Web interface. WebWasher EE can be customised easily to meet the needs of specific users. Users can create custom filter lists and criteria, including file size or character strings, and exception rules.

7.3.1 TECHNIQUES

The techniques used in the WebWasher have been developed in co-operation with the Cobion AG, a company that created a new image recognition technology. With the help of Cobion's image recognition technology WebWasher features the largest most comprehensive URL-filter database in the world - according to the information provided by the WebWasher.com - called DynaBLocator. The DynaBLocator contains a collection of URLs with the option to be blocked or not to be blocked. This database is generated in a dynamic process by webwasher.com.

The Cobion's image recognition technology is a partly automatic procedure. Cobion is permanently searching the Internet for certain image patterns and URLs that contain suspect images. Cobion sends the websites classified by the image content to the specialists at webwasher.com who consolidate and organise them in 59 categories. The database of Cobion worked out by the Webwasher-team is automatically uploaded to WebWasher EE running on a designated corporate server to integrate it in the DynaBLocator.

The pictures found are automatically analysed by a mathematical procedure of pattern recognition and classified. All by Cobion pre-classified URLs will be transformed in XML-format and sent to the WebWasher-team who control - and correct - the automatic classification of the URLs and compare the Cobion's database with databases deriving from other sources. In the end all relevant URLs will be added to one of the 59 categories of the DynaBLocator.

These categories are not all of blocking type, as the user is able to customise his/her system with which categories should be allowed and which should not. The DynaBLocator contains about ten million classified URLs. This sum increases every day as the database is daily updated. URLs can be removed if they are not longer valid. When a user tries to get access to the Internet typing in a URL, this URL will be compared with the list of the DynaBLocator. If the URL is part of the blocking categories and not released by the network administrator the access to the website is restricted.

7.3.2 EVALUATION

Although the WebWasher-team is collecting new URLs daily, it is not possible to detect all existing URLs with harmful content. A URL not included in the DynaBLocator cannot be blocked. The DynaBLocator allows an automatic research of critical image material but this pre-classification needs a lot of manual improvement and control.

An automatic text classifier does not exist in the WebWasher system, it uses a keyword search looking for strings and sub-strings in a URL. Fundamentally, the WebWasher is based on the simple technique of collecting permanently all available URLs of important content and needs a great number of human resources. The image recognition analysis is not applicable to websites, which do not use pictures or graphics; these have to be manually analysed.

The language problem has been resolved by creating regional lists that can be combined together but to cover all relevant languages of the Internet will take a few years. The aim to collect all URLs of harmful content seems to be unattainable and cannot be the only solution for effective content filtering.

8 CONCLUSIONS

This report has presented the different filtering techniques and approaches used by various tools, products and prototypes. The basic idea of this report was to analyse the maturity of the various filtering techniques and approaches that could be used for developing Internet filtering solutions.

For that, the different techniques used for the development of Internet filtering solution were divided into two major groups:

- Text-based content analysis techniques;
- And, image analysis techniques.

As it was explained previously in the report, the first group included all of these techniques and approaches that use text based content analysis. The range of solutions within this group is rather wide comprising from products which block access when certain words appear in the page textual content, to techniques which carry out semantic analysis of the text and content recognition.

Image analysis techniques include all of the solutions that classify the type of web page content, based on the analysis of the images contained in the page.

Furthermore, the consortium located various tools, products or prototypes using text-based content analysis technique and tools, products or prototypes using image analysis techniques, as well as a few tools, systems and prototypes that use both image and text-based content analysis techniques. A review of these tools, products or prototypes was conducted and the criteria for their analysis were established.

It was seen in the analysis that for the tools, products and prototypes that use only text-based content analysis techniques the results obtained are very similar to the results of those systems that use only positive and negative lists. Nevertheless, first results show that the combination of text-based content analysis techniques with positive and negative lists might improve considerably the performance of filtering solutions.

Looking at the issue of image analysis techniques combined with text-based content analysis techniques, again the results are not better than those obtained using COTS based on positive and negative lists.

Regarding the systems that only carry out image analysis techniques, they have sometimes difficulties to distinguish between what is just skin and what is nudity and much less what is pornography. Again, a good way to improve the systems available presently seems to be the combination of the different techniques with the system of lists.

Further work will be conducted in order to determine whether systems using different techniques blocked the same URLs or whether they blocked different URLs. This work will be carried out in D3.2 – Report on COTS to Integrate [3]. If this is the case, one might conclude that the combination of these techniques in these systems is actually complementary. This study is being carried out within Work Package 3 – Prototype of a European Filtering Tool.

9 REFERENCES

[1] NetProtect, NETPROTECT:WP2:D2.2:2001, "*D2.2, Report on currently available COTS filtering tools*", October 2001.

[2] GOLLER Christoph, Joachim Loning, Thilo Will & Werner Wolff, *"Automatic Document Classification: A Thorough Evaluation of Various Methods"*, ISI2000 (International Symposium on Information Sciences.

[3] NetProtect, NETPROTECT:WP3:D3.2 2001, "*D3.2, Report on COTS to Integrate*", to be published in October 2001.

[4] James Z. Wang, Jia Li, Gio Wiederhold, Oscar Firschein, "*System for screening objectionable images*", Computer Communications, vol. 21, no. 15, pp. 1355-1360, Elsevier, 1998

[5] James Z. Wang, Gio Wiederhold, Oscar Firschein, Sha Xin Wei, "*Content-based image indexing and searching using Daubechies' wavelets*", International Journal of Digital Libraries(IJODL), vol. 1, no. 4, pp. 311-328, Springer-Verlag, 1998.

10 POINTS OF CONTACT FOR FURTHER INFORMATION

Name	Role	Address/E-mail
Stéphan BRUNESSAUX	Project Director	MATRA Systèmes & Information BP 613 F-27106 Val-de-Reuil Cedex sbrunessaux@matra-ms2i.fr
Sylvie BRUNESSAUX	Workpackage Manager	MATRA Systèmes & Information BP 613 F-27106 Val de Reuil Cedex smbrunessaux@matra-ms2i.fr
Ana Luisa ROTTA	Deliverable Leader	OPTENET Paseo Mikeletegi, 58 – 1ª planta Parque Tecnológico Miramón – Edificio B8 2009 San Sebastián alrotta@optenet.com