



# **POESIA: Public Open-source Environment for a Safer Internet Access**

## **Evaluation of POESIA Beta Release**

**Sara Carro Martínez**  
**Telefónica I+D**

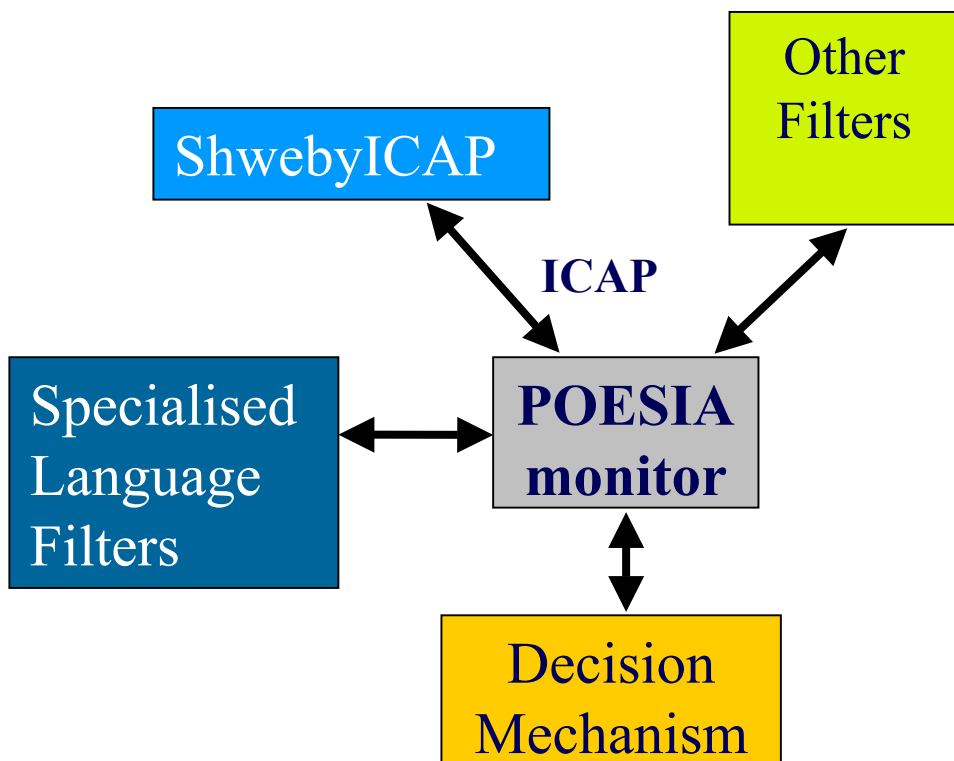
**Presented by**  
**Steve Presland**  
**Liverpool Hope**

# Evaluation of POESIA Alpha and Beta Release: Purpose

- To show the steps we have followed during the installation and compilation
- To test the system in order to produce a Final version free of errors
- To help to understand how POESIA works
- To show the future improvements that can be performance on POESIA
- To test the POESIA behaviour in different scenarios.

# Brief Overview of System Structure

- POESIA is organized around a central monitor, which receives web pages and pre-processes and distributes the content to be filtered to specialized filters and to the decision mechanism.



## Main POESIA Modules

- Monitor
- Dehtml Filter and the Language Identifier
- Image Filter
- Text Filters for English, Spanish and Italian
- URL and JavaScript Filter
- PICS (Platform for Internet Content Selection)
- Decision Mechanism
- The default Filter

# POESIA System Overview: Functionality

- **Categories of contents filtered:**
  - Pornography – Very good
  - Gross language – Good
  - Racism & Violence – Poor
- **Protocols supported:**
  - HTTP – Hyper Text Transfer Protocol.
- **Languages supported**
  - English
  - Italian
  - Spanish
  - French – to demonstrate the portability of the system.
- **Technologies**
  - URL Filtering (Black and White Lists)
  - Statistical Text filtering
  - NLP Text filtering
  - Image filtering
  - Simple JavaScript filtering
  - PICS Filter File

# Evaluation of POESIA Beta Release

- The individual modules have been tested both independently and combined together as the beta version of POESIA.
- The methodology followed for testing the integrated system was based on almost daily communications between the project team (developers and end-users), reporting the progress of the testing work, as well as the new discoveries related to POESIA.

# Testing POESIA

- **Testing Individual Filters:**
  - Initial quantitative testing of the system.
  - Each filter has been rigorously tested using a number of both non-pornographic and pornographic pages sampled from the World Wide Web.
  - The results for racism & violence filtering (in terms of symbol detection) have also been assessed.
  - The results has been used to assess the effectiveness of the techniques adopted in POESIA at using the different types of information in a page (i.e. language specific text, images, links, etc.) to filter harmful content.
- **Testing Filtering System:**
  - Complete filtering system has mainly be tested using non-pornographic and pornographic pages with a variety of content.
  - The results for racism & violence filtering (in terms of symbol detection) have also be assessed.

# Data Collection

- **Individual filters**

- **Language Identifier:** 4 116 files of approximately 200 characters
- **Text Filters:**
  - **English Filter:** 9 928 harmful and harmless pages.
  - **Italian Filter:** 7 697 harmful and harmless pages.
  - **Spanish Filter:** 4 824 harmful and harmless pages.
- **Image Filter:** 2 480 harmful and harmless images and symbols.

- **POESIA Filtering System**

- **Text Filters:**
  - **English Web pages:** 15 000 harmful and harmless Web pages.
  - **Italian Web pages:** 15 000 harmful and harmless Web pages.
  - **Spanish Web pages:** 15 000 harmful and harmless Web pages.
- **Image Filter:** 15 000 harmful and harmless images and symbols.
- **All Filters:** 60 000 files with harmful and harmless content.

# Test Scenario for the Filtering system

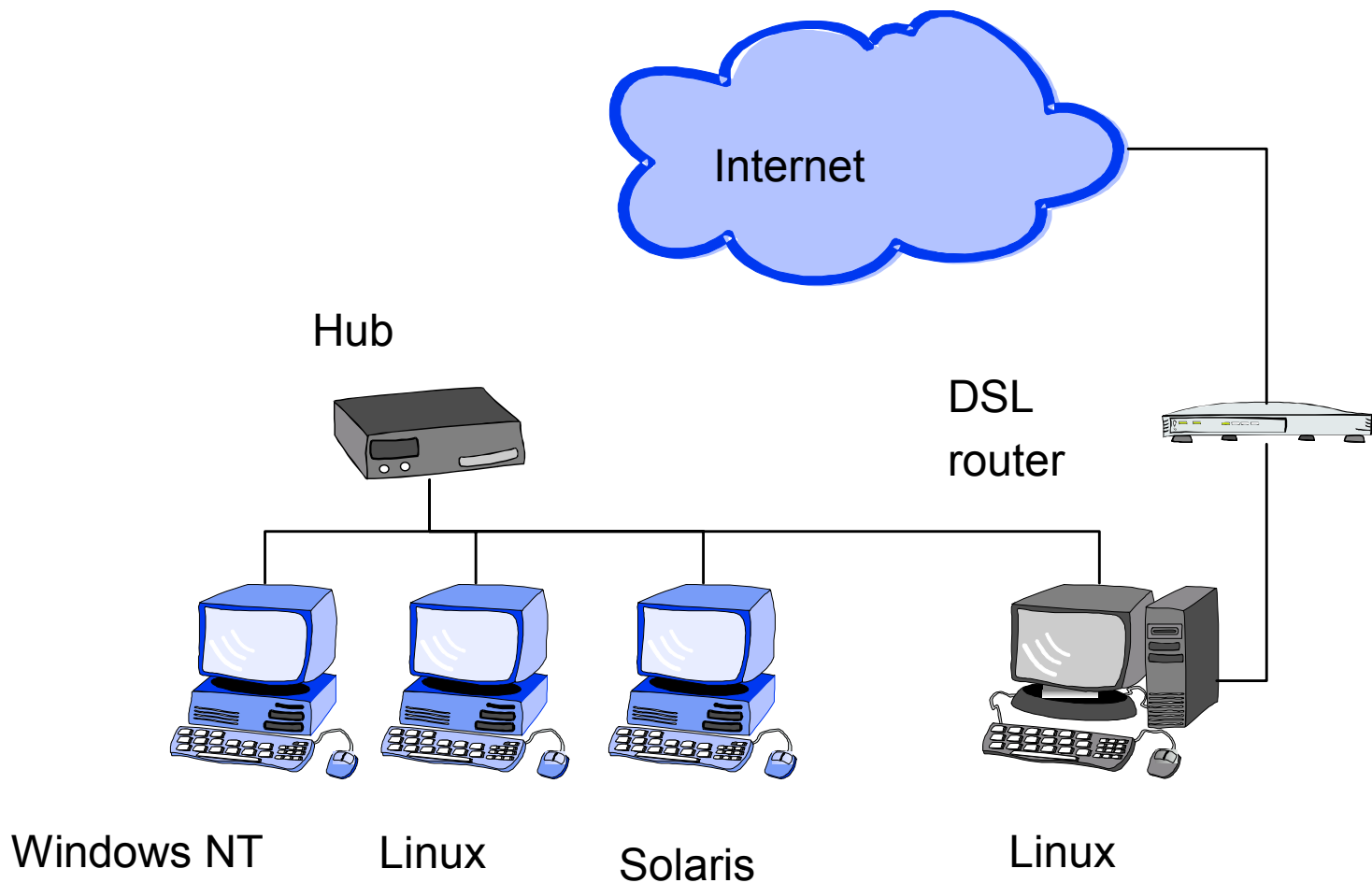
- A LAN with four user machines and a POESIA proxy.
- A different operating system was installed onto each machine:
  - The POESIA proxy with Slackware 8.1 Linux installed.

The POESIA proxy was a Pentium 4 at 2.2 GHz with 512 Mbytes of DDR RAM memory and the performance given by SPEC2000 is of 864 in integer operations and about 855 for float operations (filtering involves float operations).

- A Windows 98 machine.
- A Windows NT 4.0 machine.
- A Slackware 8.1 Linux machine.
- A Sparc Sun Solaris machine.



# Test Scenario for the Filtering system



# Results of the Evaluation: Quantitative evaluation: English Filter

## Results using English Light (Statistical) Filter

Predicted Actual	Harmful	Harmless	Unknown	Total
Harmful	4769	269	48	5086
Harmless	154	4455	233	4842
Total	4923	4724	281	9928
Precision	0.969	0.943		
Recall	0.938	0.920		
F-Measure	0.953	0.931		

## Results using English Light (Statistical) and Heavy (NLP) Filter

Predicted Actual	Harmful	Harmless	Unknown	Total
Harmful	4843	195	48	5086
Harmless	163	4446	233	4842
Total	5006	4641	281	9928
Precision	0.967	0.958		
Recall	0.952	0.918		
F-Measure	0.960	0.938		

- The addition of **the heavy filter improves the Recall** associated with harmful pages without significant adverse effect upon the Recall of harmless pages.
- Using the **combination of both filters the effectiveness increases from 0.938 to 0.952** (i.e. a reduction in the acceptance of harmful pages of nearly 25%) whilst over-blocking is only increased from 0.08 to 0.082.
- If the pages predicted as Unknown are allocated to the Harmless prediction, then the over-blocking value falls to 0.034.

# Results of the Evaluation: Quantitative evaluation: Italian Filter

## Results using Italian Light Filter

Predicted Actual	Harmful	Harmless	Unknown	Total
Harmful	3143	131	228	3502
Harmless	6	4010	179	4195
Total	3149	4141	407	7697
Precision	0.998	0.968		
Recall	0.897	0.956		
F-Measure	0.948	0.962		

## Results using Italian Light and Heavy Filters

Predicted Actual	Harmful	Harmless	Unknown	Total
Harmful	3181	165	156	3502
Harmless	15	4111	69	4195
Total	3196	4276	225	7697
Precision	0.995	0.961		
Recall	0.908	0.980		
F-Measure	0.952	0.970		

- The heavy filter has the beneficial effect of improving the overall classification by reducing the proportion of pages classified as Unknown.
- The addition of the heavy filter does not decrease the rate of Harmful pages misclassified as Harmless.
- **The heavy filter improves the Recall associated with both pornographic and non-pornographic pages without significantly affecting the Precision values.**

# Results of the Evaluation: Quantitative evaluation: Spanish Filter

## Results using Spanish Light (Statistical) Filter

Predicted Actual	Harmful	Harmless	Total
Harmful	816	75	891
Harmless	4	3929	3933
Total	820	4004	4824
Precision	0.995	0.981	
Recall	0.916	0.999	
F-Measure	0.954	0.990	

- The classification technique employed by the Spanish filter allocates the pages predicted as Unknown to the harmless category.
- From the table it can be seen that the **effectiveness value of the filter is 0.916** whilst the **over-blocking value is only 0.001**.

# Results of the Evaluation: Quantitative evaluation: Image Filter

## Pornographic detection: Results using symbol filter

Predicted Actual	Harmful	Harmless	Total
Harmful	910	90	1000
Harmless	200	800	1000
Total	1110	890	2000
Precision	0.82	0.90	
Recall	0.91	0.80	
F-Measure	0.86	0.85	

## Harmful symbol detection: Results using symbol filter

Predicted Actual	Harmful	Harmless	Total
Harmful	850	150	1000
Harmless	110	890	1000
Total	960	1040	480
Precision	0.885	0.856	
Recall	0.85	0.89	
F-Measure	0.867	0.873	

- The image filter for the identification of pornographic images provides an effectiveness of 0.91 with a over-blocking value of 0.2.
- **The image filter is only categorising a single image rather** than all the image content found on a given web page.

- **Very low elapsed time compared with traditional filters.**
- Difficulty of the symbol detection domain.

# POESIA Filtering System : Decision Mechanism

- The Decision Mechanism generates a decision based on a configurable strategy that takes into account the responses of all filters.
- For the evaluation of the Beta version the Decision Mechanism used a strategy whereby if any filter returned a “high” score, or half or more of the filters returned a “medium” score then the page was blocked, otherwise the page was allowed.

# Effectiveness and Over-blocking

Effectiveness:

$$\frac{\text{Number of harmful pages blocked}}{\text{Total number of harmful pages}}$$

Over-blocking:

$$\frac{\text{Number of harmless pages blocked}}{\text{Total number of harmless pages}}$$

# Initial Evaluation: Text Filters

## Effectiveness

## Over-blocking

TEXT FILTER	INDIVIDUAL FILTER	POESIA	INDIVIDUAL FILTER	POESIA
English	0.952	0.969	0.082	0.063
Italian	0.908	0.940	0.020	0.050
Spanish	0.916	0.973	0.001	0.028

Comparing the results for the POESIA system with those for the individual text filters in isolation it can be seen that there is:

- An overall improvement in effectiveness
- Varied impact in the over-blocking results



# Initial Evaluation: Image Filter

## Effectiveness

## Over-blocking

<b>FILTER</b>	<b>INDIVIDUAL</b>	<b>POESIA</b>	<b>INDIVIDUAL</b>	<b>POESIA</b>
<b>Image</b>	<b>0.910</b>	<b>0.933</b>	<b>0.200</b>	<b>0.068</b>

Comparing the results for the POESIA system with those for the individual image filter in isolation it can be seen that there is an improvement in both :

- effectiveness, and
- over-blocking

Many image-only Web pages are likely to contain multiple images, therefore the image filter has more information upon which to make the classification and thus improve its accuracy.

# POESIA Filtering System: All Filters

## Results using the POESIA filtering system on all web pages

The POESIA filtering system provides an effectiveness value of 0.954 with an over-blocking value of 0.028.

The poorest performance is seen on the image only web pages and by excluding these the effectiveness value increases to 0.961, and the over-blocking value decrease to 0.023.

- **NetProtect project** evaluation:
  - effectiveness value of 0.957 with an over-blocking value of 0.165.
  - effectiveness value of 0.884 with an over-blocking value of 0.020.
- In one reported filtering test (available at [http://www.veritest.com/clients/reports/websense/websense\\_really.pdf](http://www.veritest.com/clients/reports/websense/websense_really.pdf)) the system provides an effectiveness value of 0.95 with virtually no over-blocking,
- A direct comparison of these results is difficult due to differences in the number of pages tested:
  - the NetProtect project evaluation used only 3114 pages.
- The second project only used a small test set (200 pornographic pages) were acquired by placing specific terms into a search engine, thus guaranteeing that the pages contain terms on which the system can filter effectively.

# Comparison of Results

<b>Filtering System</b>	<b>Effectiveness</b>	<b>Over-blocking</b>
<b>POESIA</b>	<b>0.954</b>	<b>0.028</b>
<b>POESIA (removing image only pages)</b>	<b>0.961</b>	<b>0.023</b>
<b>NetProtect (with OR formula)</b>	<b>0.957</b>	<b>0.165</b>
<b>NetProtect (with MV formula)</b>	<b>0.884</b>	<b>0.020</b>
<b> Websense<sup>1</sup></b>	<b>0.950</b>	<b>Virtually 0</b>

(<sup>1</sup>report available at [http://www.veritest.com/clients/reports/websense/websense\\_really.pdf](http://www.veritest.com/clients/reports/websense/websense_really.pdf))

# Qualitative criteria

- Usability – Friendliness
- Usability – Understandability
- Operational integrity
- Unblocking service
- Configurability
- Other features
- Cost

# User-friendliness

- **Monitor**

- Hard – expert knowledge required for installation
- Mainly manual installation
- Requires reconfiguration of client browser
- Takes a long time to install/uninstall

- **Individual filters**

- Easy to moderate knowledge required
- Generally semi-automatic installation
- Does not require configuration of browser
- Takes a short while to install/uninstall

# Qualitative evaluation: Monitor

<b>Usability – Friendliness</b>	
How Easy was the installation?	Expert PC knowledge (hard)
Was the installation completely automatic, semi-automatic or did it require great manual procedure?	Manual Procedure
Did the installation require configuration of the browser?	Yes
How long did the installation take?	Much Time
How long did it take to remove the filter?	Much Time (with all the packages)
How easily can the filtering software be removed?	Expert PC knowledge (hard)
Did you have to reboot your system?	No
Did you have to reinstall your operating system?	No
<b>Usability – Understandability</b>	
Is there a trace (log file) kept of the monitor activity?	Yes, many traces
Are log files analysable and printable?	Yes
<b>Protocols supported</b>	
HTTP	Yes
ICAP	Yes
<b>Technologies applied (urlists)</b>	
White Lists	Yes
Black lists	Yes
<b>Unblocking service</b>	
Does it provide an unblocking service for Web pages blocked by mistake?	No

# Qualitative evaluation: Language Identification Filter

<b>Usability - Friendliness</b>	
How Easy was the installation?	Easy/Moderate
Was the installation completely automatic, semi-automatic or did it require great manual procedure?	Semi-automatic
Did the installation require configuration of the browser?	No
How long did the installation take?	A few minutes
How long did it take to remove the filter?	A few minutes
How easily can the filtering software be removed?	Easy
Did you have to reboot your system?	No
Did you have to reinstall your operating system?	No
<b>Usability – Understandability</b>	
Is there a trace (log file) kept of the filter activity?	Yes
Are log files analysable and printable?	Yes
<b>Languages supported</b>	
English; French; Italian; Spanish; Danish; Dutch; Finnish, German.Greek, Portuguese, Swedish	Yes
<b>Technologies applied</b>	
Content analysis	Yes

# Qualitative evaluation: The English Filter

<b>Usability - Friendliness</b>	
How Easy was the installation?	Easy-Moderate
Was the installation completely automatic, semi-automatic or did it require great manual procedure?	Semi-automatic
Did the installation require configuration of the browser?	No
How long did the installation take?	A few minutes
How long did it take to remove the filter?	A few minutes
How easily can the filtering software be removed?	Easy-Moderate
Did you have to reboot your system?	No
Did you have to reinstall your operating system?	No
<b>Usability – Understandability</b>	
Is there a trace (log file) kept of the filter activity?	Yes
Are log files analysable and printable?	Yes
<b>Categories of contents filtered</b>	
Pornography	Yes
Gross Language	Yes
<b>Technologies applied</b>	
Text / Keyword filtering	Yes
Content analysis	Yes



# Qualitative evaluation: The Spanish Filter

<b>Usability – Friendliness</b>	
How Easy was the installation?	Easy-Moderate
Was the installation completely automatic, semi-automatic or did it require great manual procedure?	Semi-automatic
Did the installation require configuration of the browser?	No
How long did the installation take?	A few minutes
How long did it take to remove the filter?	A few minutes
How easily can the filtering software be removed?	Easy-Moderate
Did you have to reboot your system?	No
Did you have to reinstall your operating system?	No
<b>Usability – Understandability</b>	
Is there a trace (log file) kept of the filter activity?	Yes
Are log files analysable and printable?	Yes
<b>Categories of contents filtered</b>	
Pornography	Yes
Gross Language	Yes
<b>Technologies applied</b>	
Text / Keyword filtering	Yes
Content analysis	Yes

# Qualitative evaluation: The Italian Filter

<b>Usability – Friendliness</b>	
How Easy was the installation?	Moderate
Was the installation completely automatic, semi-automatic or did it require great manual procedure?	Semi-automatic/Manual
Did the installation require configuration of the browser?	No
How long did the installation take?	Moderate
How long did it take to remove the filter?	Moderate
How easily can the filtering software be removed?	Moderate
Did you have to reboot your system?	No
Did you have to reinstall your operating system?	No
<b>Usability – Understandability</b>	
Is there a trace (log file) kept of the filter activity?	Yes
Are log files analysable and printable?	Yes
<b>Categories of contents filtered</b>	
Pornography	Yes
Gross Language	Yes
<b>Technologies applied</b>	
Text / Keyword filtering	Yes
Content analysis	Yes

# Qualitative evaluation: The Image Filter

<b>Usability – Friendliness</b>	
How Easy was the installation?	Moderate
Was the installation completely automatic, semi-automatic or did it require great manual procedure?	Semi-automatic
Did the installation require configuration of the browser?	No
How long did the installation take?	Moderate time
How long did it take to remove the filter?	Moderate time
How easily can the filtering software be removed?	Moderate
Did you have to reboot your system?	No
Did you have to reinstall your operating system?	No
<b>Usability – Understandability</b>	
Is there a trace (log file) kept of the filter activity?	Yes
Are log files analysable and printable?	Yes
<b>Categories of contents filtered</b>	
Pornography	Yes
Symbol	Yes

# Qualitative evaluation: The URL and PICS Filter

<b>Usability – Friendliness</b>	
How Easy was the installation?	Easy
Was the installation completely automatic, semi-automatic or did it require great manual procedure?	Semi-automatic
Did the installation require configuration of the browser?	No
How long did the installation take?	A few minutes
How long did it take to remove the filter?	A few minutes
How easily can the filtering software be removed?	Easy
Did you have to reboot your system?	No
Did you have to reinstall your operating system?	No
<b>Usability – Understandability</b>	
Is there a trace (log file) kept of the filter activity?	Yes
Are log files analysable and printable?	Yes
<b>Categories of contents filtered</b>	
Pornography	Yes
Gross Language	Yes
<b>Technologies applied</b>	
Black lists	Yes (URL)
Rules-based content analysis	Yes (PICS)
<b>Protocols supported</b>	
PICS	Yes (PICS)

## Monitor

Operational integrity	Small	Medium	Large
How much does it slow the Internet traffic? (0-3)	1	1	3
Does it interfere with other applications?	No	No	Yes
Does it have a stable behaviour (ie. No crash problem)?	Yes	Yes	Yes

## Language Identifier, English Filter, Spanish Filter

Operational integrity	Small	Medium	Large
How much does it slow the Internet traffic? (0-3)	0-1	0-1	1-2
Does it interfere with other applications?	No	No	No
Does it have a stable behaviour (ie. No crash problem)?	Yes	Yes	Yes

## Italian Filter

Operational integrity	Small	Medium	Large
How much does it slow the Internet traffic? (0-3)	1	1	2
Does it interfere with other applications?	No	No	No
Does it have a stable behaviour (i.e. No crash problem)?	Yes	Yes	Yes

## Image Filters

Operational integrity	Small/Medium/Large
How much does it slow the Internet traffic? (0-3)	2
Does it interfere with other applications?	No
Does it have a stable behaviour (ie. No crash problem)?	Yes

## URL & PICS Filters

Operational integrity	Small/Medium/Large
How much does it slow the Internet traffic? (0-3)	Not pertinent
Does it interfere with other applications?	No
Does it have a stable behaviour (ie. No crash problem)?	Yes

# Possible Future Developments - 1

- The architecture of the **POESIA system is highly modular**, and has well-specified interfaces which **facilitates outside contributions to the system**. i.e. Shewby proxy.
- The most obvious improvement would be **to add filters** to enable the filtering of other languages.
- Using the current implementation, it would be possible to **extend the filtering to other domains**.  
The constraint: collection of training (and testing) data.
- Filters could also be added **to handle other content types**, such as PDF, Word files, Excel files, etc.
- **other forms of Internet communication**, such as Chat, News and E-mail could be handled. These protocols are also not limited to the interchange of words but also the exchange of images, video, etc.
- Accommodation of other internet access technologies, e.g. WAP

## Possible Future Developments - 2

In terms of extending the functionality of the current software:

- **a mechanism to ask and grant permission** could be developed.
- **dynamic white and black lists** could be included.
- **content from other sources** such as CD-ROM, Floppy disk, etc could be analysed.

# Conclusions

- The performance of POESIA has been evaluated on two levels:
  - Quantitatively:
    - The filtering approach produces comparable results with other filtering system with an overall effectiveness value of 0.954 and over-blocking value of 0.028.
  - Qualitative evaluation:
    - The system proved to be more complex to implement and run than originally envisaged, however it should be seen in the context of a complex OpenSource system.
    - The installation process could be improved, for example with the use of an RPM package.



# Conclusions

- POESIA implements an effective state-of-the-art content filtering system combining text and image filters, as well as the more commonplace URL, JavaScript and PICS filters.
- The OpenSource nature of the project will allow future developments and improvements.