POESIA: *P*ublic *O*pen-source *E*nvironment for a *S*afer *I*nternet *A*ccess

POESIA Final Report
Deliverable 1.4

Project name:	POESIA
Project number:	IAP 2117/27572
Date:	April 3rd 2004
Document Id:	POESIA-WP1-1.4
Version:	1.0
Deliverable:	1.4
Related WP:	1 - Project Management and Coordination
Authors:	ILC with the contribution of all consortium partners
Status:	Public

Executive summary

This document is the final report of the POESIA project. It describes the objectives of the project, the work performed and the results achieved during the project lifetime. It also envisages a range of possible improvements and extensions which might be desirable for future versions of the POESIA filtering system.

The overall goal of the project was to develop, test, evaluate and promote a fully open-source, extensible, state of the art, filtering and caching software solution, in order to achieve more effective filtering than currently provided by existing products. The POESIA system aimed at filtering several channels by combining highly innovative technologies. A key feature of the project was its aim of providing advanced content-based filtering, both on the basis of textual and image content. Filtering in POESIA was thus expected to cover a range of modes, including image filtering, natural language text filtering (for English, Italian, Spanish and – at a later stage – for French), URL, PICS and JavaScript filtering. All objectives foreseen in the POESIA Technical Annex have been successfully achieved in due course.

The main results of the POESIA project can be summarized as follows:

1. a successful open-source filtering system resulting from the integration of multiple filters, each of which addressing some source of evidence that is of potential use in identifying harmful content. The POESIA system combines components based on standard filtering methods, such as positive/negative URL lists and PICS, with components incorporating more advanced techniques, such as image processing and NLP-enhanced text filtering. The evidence detected by individual filters is then combined by a Decision Mechanism (DM) component to produce an overall decision for each page. The final version of the POESIA system focuses on filtering harmful content from the pornographic domain, with some more limited attention being given to the filtering of gross language (in text) and the identification of material having violent/hate associations (through the recognition in images of symbols that are logos for relevant groups). The languages covered are English, Italian, Spanish; a partial, limited, proof of concept, port to French has been done with encouraging results. The system has been extensively tested against various sets of URLs, both in controlled and live environments, using different scenarios and configurations, and always showed a performance in line with other state-of-the-art filtering systems;
2. the publication of several project reports, available for download on a dedicated dissemination website (see below), and of numerous papers presented at international conferences and workshops;
3. dedicated Websites for dissemination (http://www.poesia-filter.org/index.htm) and OpenSource development (http://sourceforge.net/projects/poesia/) where project documentation and software are available.

The Open Source nature of the software combined together with the architectural characteristics of the system provides a basis for development and maintenance of the POESIA system beyond the project's lifetime: there are many conceivable extensions of functionality and operation that would constitute valuable additions to the system which have been delivered among the project outcomes.

Table of Contents

1 Introduction

This document is the final report of the POESIA project. It describes the objectives of the project, the work performed and the results achieved during the project lifetime. It also envisages a range of possible improvements and extensions which might be desirable for future versions of the POESIA filtering system.

2 Project overview

2.1 *The Safer Internet Action Plan*

POESIA is a project partially funded by the European Commission and takes place in the mark of the Information Society and Technology Safer Internet Action Plan. The Safer Internet Action Plan (IAP) promotes safety on the Internet by tackling the controversial issue of illegal, harmful and racist content. It supports a safer Internet through three main action lines. The POESIA project fits inside the second line: development of rating and filtering systems for Internet content. The project officially started on February 4th 2002 and lasted for two years.

2.2 *Project goals*

The overall goal of the project was to develop, test, evaluate and promote a fully open-source, extensible, state of the art, filtering and caching software solution, in order to achieve more effective filtering than currently provided by existing products. The POESIA system aimed at filtering several channels by combining highly innovative technologies. A key feature of the project was its aim of providing advanced content-based filtering, both on the basis of textual and image content. Filtering in POESIA was thus expected to cover a range of modes, including image filtering, natural language text filtering (for English, Italian, Spanish and – at a later stage – for French), URL, PICS and JavaScript filtering. All these objectives have been successfully achieved.

2.3 *POESIA as an open-source project*

The name of the project, which stands for Public Open-source Environment for a Safer Internet Access, highlights a key feature of the project, i.e. that POESIA is developed as an open-source (see http://www.opensource.org/) or free[1] filtering software system. This entails that every piece of software developed under the POESIA project is released under a free software licence (usually the GPL or LGPL, see http://www.gnu.org/copyleft/), in source code form. This source code is viewable, reusable and improvable by peer developers, and should be compiled into executables to produce working filtering software. Free availability of the POESIA system as an open-source software will provide numerous market opportunities to commercial corporations (both large and SME), in particular in network configuration and administration businesses, educational businesses, etc. It will also permit a wide deployment of the filtering system in many educational institutions. Last but not least, the fact of being Open-Source provides a basis for development and maintenance of the system beyond the project's lifetime (see section 4.4 below).

[1] We use both terms interchangeably -even if they are significant philosophical differences-, but we are more in the free (or libre) software side, since POESIA uses mostly the GPL or LGPL free software licenses.

2.4 *POESIA Target Audience*

The main targets of POESIA are educational institutions (Internet classrooms in primary or secondary schools, colleges, universities) and other collectivities (e.g. libraries, cybercafés, museums, corporate networks) in need of Internet content filtering, mostly with young users. In particular, the POESIA target audience was identified in decision makers (e.g. heads of organisations, IT co-ordinators and network managers), as these users are a key audience in determining the likely use of Internet filtering options.

2.5 *Project results*

The main achievements of the POESIA project can be summarized as follows:

4. a successful open-source filtering system resulting from the integration of multiple filters, each of which addressing some source of evidence that is of potential use in identifying harmful content. The POESIA system combines components based on standard filtering methods, such as positive/negative URL lists and PICS, with components incorporating more advanced techniques, such as image processing and NLP-enhanced text filtering. The evidence detected by individual filters is then combined by a Decision Mechanism (DM) component to produce an overall decision for each page. The final version of the POESIA system focusses on filtering harmful content from the pornographic domain, with some more limited attention being given to the filtering of gross language (in text) and the identification of material having violent/hate associations (through the recognition in images of symbols that are logos for relevant groups). The languages covered are English, Italian, Spanish; a partial, limited, proof of concept, port to French has been done with encouraging results. The system has been tested against various sets of URLs and always showed a performance in line with other state-of-the-art filtering systems;
5. the publication of several project reports (listed in Appendix 1), available for download on a dedicated website (see below), and of numerous papers presented at international conferences and workshops;
6. dedicated Websites for dissemination and OpenSource development where project documentation and software are available.

3 The Consortium

3.1 *The project partners*

The project was carried out by a wide consortium of 10 members listed in the table below:

Participant Short Name	Participant Name	Country
CEA	Commissariat à l'Energie Atomique	Gif/Yvette, FRANCE
ENIC	Ecole Nouvelle d'Ingénieurs en Communication	Villeneuve-d'Ascq, France
FCR	Fundació Catalana per a la Recerca	Barcelona, SPAIN
HOPE	Liverpool Hope University College	Liverpool, UNITED KINGDOM
ILC	Istituto di Linguistica Computazionale	Pisa, ITALY
META	M.E.T.A. S.r.l.	Lucca, ITALY

Participant Short Name	Participant Name	Country
PIXEL	PIXEL Associazione	Firenze, ITALY
T, I+D	Telefónica Investigación y Desarrollo S.A. Unipersonal	Boecillo, SPAIN
UEM	Universidad Europea de Madrid	Villaviciosa de Odón, SPAIN
USFD	University of Sheffield	Sheffield, UNITED KINGDOM

3.2 *Organisation*

The POESIA consortium was jointly led by CEA for what concerns technical management and by ILC for administrative and financial coordination.

The POESIA consortium was organised into two sub-groups which actively and continuously interacted throughout the project:

- the "end users" group, coordinated by HOPE and including FCR, PIXEL and T, I+D, that defined detailed end-users requirements, provided test case examples to evaluate the filtering software, assessed usability and performance of the POESIA system and promoted a wide-ranging dissemination for the project activities and outcomes;
- the "developers" group, coordinated by CEA and including ENIC, ILC, META, UEM and USFD, that designed the software architecture and carried out its implementation.

Concerning internal project communications, a number of systems have been implemented for all partners to participate in the project development. These include group email and different mailing lists, virtual platforms (CVS and Learnwise), regular project meetings and ftp and sftp sites. All partners in the Project team have contributed to decisions made and to the activities undertaken, as described in the contract and work plan.

4 Main results achieved

4.1 *From End User Requirements to POESIA Recommendations*

The End User team of the POESIA project surveyed current and on-going activities in the area of Internet filtering, explored End User opinions and established needs and recommendations for the POESIA technical development team. End Users (in particular, decision makers, advisors and experts in educational establishments) were surveyed between June and July 2002, mainly from the three participating countries of the End User group (Italy, Spain and the UK). A wider audience has been sought through an on-line survey and questionnaires distributed at European conferences and workshops. The results of this survey are based on 261 filled questionnaires distributed as follows: Spain (130); UK (87); others, i.e. mainly Italy, (44). The research demonstrated that while there are similarities in the overall concerns and perceived needs of many educational organisations, there were some significant cross-cultural differences in their approaches and attitudes to the safer Internet options that they provide in their schools, colleges, universities and libraries across Europe. The updated findings of this survey are documented in the public deliverable D2.1 "End-user Requirements Report" and have also been disseminated through several academic presentations and conference papers.

The general requirements proposed by the end users team for the design of the POESIA system were carefully discussed and evaluated with the developer partners to assess their feasibility and actual implementation in the project lifetime. This was done by taking into account a number of different factors:

- resource limitations (i.e. the duration of the POESIA project, which is only two years);
- state-of-the-art in the fields of NLP and image processing;
- performance requirements for practical filtering (e.g. acceptable waiting time, acceptable error rate).

Key decisions were made by the whole consortium at different POESIA meetings where it was agreed to focus the development effort for content-based filtering towards only certain domains and channels.

In regard to channels, the focus has been on web pages and – in a more restricted way – on email (only porn-oriented SPAM messages). Within the web pages channel, three different content categories have been addressed:

1. pornography, dealt with through NLP and Image based filtering;
2. gross language, handled by the NLP-based components;
3. violence, which is addressed in a restricted fashion, via a very innovative use of image processing techniques to identify symbols that are related to potentially harmful content, in particular symbols associated with hate groups.

Obviously, illegal content (such as paedophilic pornography) was outside the realm of POESIA because gathering illegal examples and testing with them would have required sensitive specific authorizations from law officials that the Consortium could not acquire; it is important to point out, however, that the technology developed within the project might catch part of it.

Last but not least, it should be noted that, although some domains will not be addressed by content-based filtering, alternative filtering methods such as URL-based filtering, are still available in the POESIA system to prevent access to inappropriate material.

4.2 *Design and development of the POESIA filtering system*

4.2.1 The POESIA scenario

The POESIA system was designed to run on a GNU/Linux (free software) system, usually on a PC (unattended) box[2] or a semi-dedicated (e.g. the teacher's) Linux workstation. This GNU/Linux system should have two network interfaces: one for the connection to the outside Internet and another for the internal (classroom) local network. So every bit of information has to flow through the POESIA box before it hits a browser on the internal local network (e.g. the classroom): POESIA filtering is unavoidable, and does not require any specific browser (or configuration) on the classroom browsing stations (which can be of any kind and run any software compliant with Web standards).

4.2.2 Overall architecture of the POESIA system

[2] The box running POESIA could be a barebone or rackable or brick model, without any keyboard, mouse or screen.

The POESIA filtering system is organized around a central monitor, which deals with external information sources (e.g. web, e-mail) and pre-processes and distributes the information to be filtered to the specialized filters – namely the language detector, the Natural Language Processing filters, the Image filter and the URL, PICS and JavaScript filter - and to the decision-making components. The overall architecture of the system is sketched in Figure 1:

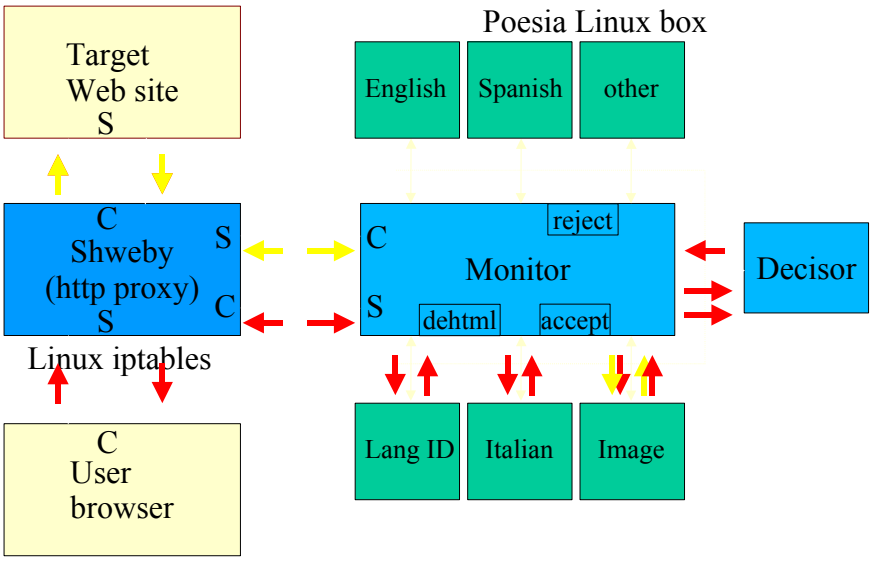

Figure 1: POESIA overall architecture

The POESIA monitor is the sole program communicating (directly) with filtering clients. Filters and decision-making components communicate directly only with the POESIA monitor. The POESIA monitor starts, schedules and communicates with POESIA specific filters (e.g. image or NLP filters) and sub-processes (like the decision mechanism). Filters can not only produce scores (which are handled by decisors) but can also ask (through a continuation message) for another filter to process their intermediate result. For instance, in this way, a language detector filter can pass textual content to a language specific (Italian, Spanish, English) filter.

The filtering system can be configured to provide different levels of filtering at different times of day, and/or for different browsing clients, etc. The POESIA architecture is documented in the publicly available "Software Architecture Definition" report (D3.1). Although the software architecture definition was supposed to end at month 8, the continuous development of the software required minor architectural changes or enhancements: in particular, there are architecture details, concerning the interaction between the monitor and the filters and the decision mechanisms, which needed to be refined while the development of these components was progressing.

The POESIA architecture was explicitly designed with configurability and extensibility in mind: the same POESIA monitor is usable to filter content types not yet filtered in this project, by addition of internal filters and reconfiguration; for more details, see D9.4 "Definition of Future Improvement Report" and D9.3 "Evaluation Report of POESIA Software Final Release" (in particular, section 5.3 "How to add a new filter to POESIA").

4.2.3 Image filtering

The POESIA system provides image filtering, in particular for filtering of Web sites. Only fixed pixelised images (GIF, JPEG, PNG, etc.) are filtered within POESIA (so MPEG movies, Macroflash images, and vector SVG drawings are out of scope). Since images are embedded in HTML, image filtering requires pre-fetching of images contained inside an HTML page (before the browser requires them).

Image filtering in POESIA is focused towards two independent goals:

1. detection of pornographic images;
2. detection of harmful symbols (like nazi crosses or some sectarian cult symbols) against a configurable set of model symbols.

The final version of image processing and filtering components (D6.2 "Skin Feature Extraction and Detection Component"), which includes the pornographic image filter and the symbol detection component, is available at http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/ImageFilter/ and http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/Java/. The work done is documented in Section 6 "Image filtering" of D3.1 "Software Architecture Definition Document", and in a number of scientific papers presented at international conferences.

Training of the image filters was carried out on the basis of both appropriate training corpora of examples specifically built by scanning the Web (see D6.1 "Image Subsets") and of the manually classified database of URLs collected by end-users (see D2.2 "Test Case Files").

4.2.3.1 Pornographic image filter

The internal architecture of the pornographic image filter includes:

1. a skin detection component and
2. a form analysis component.

For each input image, the POESIA skin detection component outputs a greyscale skin map with the grey level indicating the probability of skin. All the pixel-based skin detection techniques in the literature adopt the independent model, i.e. they classify each pixel individually, not taking into account the correlation between neighboring pixels. However, the skin pixels do not distribute purly randomly, they tend to form some regions, neighboring pixels do correlate. Thus, we proposed a model that can capture this mutual information. We build this model with maximum entropy modeling, and the model built is a Markov Random Field, which is more reasonable than the baseline independent model. According to a recent survey of the most widely used pixel-based skin detection techniques by Vezhnevets et al. (in Graphicon 2003, 13th International Conference on the Computer Graphics and Vision, Moscow, Russia, September 2003), the skin detection algorithm proposed within POESIA appears to have the best performance in terms of pixel classification rates. Skin detection permits a large number of images that do not contain skin to be excluded from the following steps of the filtering process.

However, non-pornographic images such as portraits or group portraits may also contain considerable portions of skin. Fortunately, skin regions in pornographic images have specific shapes, generally different from those of non-pornographic images such as portraits or group portraits. In order to distinguish pornographic images from non-pornographic images containing skin, the POESIA pornographic image filter analyses the shape information of skin regions in images. This task is carried out by the form analysis component, which operates on the output of the skin detection component and aims at detecting skin regions features. Several features are extracted from the skin map of the original image, which permits to give an overall description of the skin region shapes. A multi-layer perceptron classifier is then trained for these features. The training image database contains more than 5,000 images with both pornographic images and non-pornographic images. After training, the classifier has learned the difference between pornographic image features and non-pornographic image features and is ready to reliably classify any input image.

4.2.3.2 Symbol detection component

The goal of symbol detection is to compare a symbol entirely filling an image[3] against a small fixed set of model symbols. This can be not only useful for content filtering of images (e.g. detecting every occurrence of a flag or logo) but - being available in open-source form - could also be reused in other settings (e.g. for the detection of navigational arrows). The guiding principle underlying this task is to extract from a symbolic image a vector of descriptors (a tuple of real numbers) which is invariant for usual image operators like translation, scaling, limited rotation, and then to compare the descriptive vector against those of the model symbols.

From the image symbols collected by POESIA End Users, a catalogue of more than two hundred images was created and subdivided into two subsets: 194 harmful symbols and 21 non harmful symbols. These symbols were collected from different web sites (racist, xenophobic, cults [i.e. religious intolerant sects], violence). Invariant descriptors (Zernik moments) were computed for each of the symbols to be filtered out. Such descriptors are then matched against the descriptive vector representing input symbols coming from the web.

4.2.4 Natural Language Processing text filtering

The POESIA system provides textual content filtering for three European languages, namely English, Italian and Spanish, employing advanced Natural Language Processing (NLP) techniques to enhance filtering effectiveness and performance.

Within POESIA, the three language-specific text filters have been developed by different sites which specialise in NLP for the target languages, which are USFD for English, ILC for Italian and UEM for Spanish. The filters differ in the methods they employ, partly reflecting an attempt to optimise over the different NLP resources available for each language.

To assess the ease of porting the POESIA technology to another language, a partial, limited, proof of concept, port to French has been done with encouraging results. The work done for French mainly involved the re-application of language-independent subcomponents developed for the language-specific text filters, with a view to assessing the extent to which these readily portable subcomponents can provide adequate basic filtering for other languages.

The POESIA system also includes a language identifier component, which is required to ensure that page text is routed to the appropriate language-specific text filter.

The current POESIA implementation performs text filtering only for URLs whose target is a plain text page or a HTML page. The domains filtered are pornography and gross language.

The work done for the different languages is documented in Section 7 “NLP-based text filtering” of D3.1 “Software Architecture Definition Document”, in D7.2 “Lexical Resources and Tools for Each Language”, and in a number of scientific papers presented at international conferences and workshops.

The final version of the NLP-based filtering software components for all languages (including French) is available at http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/TextFilter/.

[3] Detection of a symbol as a small part of a larger image is out of scope.

4.2.4.1 Text filtering: general issues

A tag stripper removes HTML tags from HTML. Then, such textual content is first processed by a language identification filter, which redirects it (using a continuation message to the monitor) to a language specific text filter.

Language-specific text filters offer both "light" and "heavy" filtering modes, where the former is simpler and faster, and the latter is computationally more expensive. Light filtering, which uses little NLP, provides rapid accept/reject decisions for straightforwardly classifiable pages. For other pages, heavy filtering, making greater use of NLP, is invoked to provide more sensitive detection of content indicators. This trade-off is important to the overall efficiency of the system.

4.2.4.2 Data acquisition

The development and testing of the POESIA text filters required a substantial quantity of pages for each language, which have been precategorised as pornographic and non-pornographic. A first starting point was provided by end-users (D2.2 "Test Case Files"). However, the fully manual collection of this data would have been infeasible. Therefore, the data was automatically spidered from the WWW, using the Google directory structure (http://directory.google.com) to locate sites which fall into the pornographic or non-pornographic category. The spider tranverses links within identified sites to retrieve pages at varying depths. Pages are stripped of HTML and the text is analysed to ensure it is of the target language and to highlight potential misclassifications. Despite these checks, there have inevitably been some number of pages which were incorrectly classified, and this fact is reflected in the final performance scores. The corpus collected for each language, ranging in size between 5k and 20k pages, was gathered in D7.1 "Domain Corpora for Each Language".

4.2.4.3 Language identification

The languages addressed for text filtering in POESIA were English, Spanish, Italian and (at a later stage) French. Before text filtering can be done, the language used within a document must be identified. A component to perform language identification has been implemented in Java. The approach used computes statistics for letter n-grams in the supported languages, and compares these statistics to those obtained for each new document, to determine its most likely language.

4.2.4.4 Text Filtering for English

The English light filter employs a bag-of-words approach, with stoplisting and stemming, and with indexing terms selected via a minimum threshold for document frequency in the training corpus. The current implementation classifies using an out-of-place measure over term frequency rankings. The English heavy filter focusses on pages misclassified as non-pornographic by the light filter during training. A set of keywords is identified from these pages (words which are commonly indirect indicators of pornographic content, eg. adult, explicit). An instance-based approach is used to learn contextual differences for the use of these keywords in non-pornographic vs. misclassified pornographic pages. Contextual features include adjacent words, their POS tags, and named entities. At runtime, documents classified as non-pornographic by the light filter, which contain any keywords, are passed to the heavy filter. This applies the contextual patterns to determine the predicted document class.

4.2.4.5 Text Filtering for Italian

The Italian light filter employs statistical word-based categorization, using local term counts rather than global frequencies. Text is tokenised and segmented into windows of 100 words. Each window is assigned a score based on the maximum local frequency of domain relevant words (markers). For each text, the filter outputs the maximum cumulative word score over different text windows. For efficiency, given the morphological richness of Italian, the morphological variants of ~40 unambiguous marker lemmata, extracted from a linguistically annnotated training courpus, were precomputed. Thresholds map text scores to low/medium/high values. Low/high values are notified directly to the monitor to be passed to the DM component. For medium values, the heavy filter is invoked. This filter operates on morpho-syntactically tagged and lemmatized texts and is based on recognition of ~2400 domain relevant lemmata (including ambiguous words). Category learning uses an entropy-based classifier: CASSANDRA (Complex Analysis of Sequences via Scaling AND Randomness Assessment), which computes the rate of information increase generated by salient lemmata.

4.2.4.6 Text Filtering for Spanish

The Spanish light filter uses state-of-the-art text categorization techniques. Pages are firstly tokenized, stoplisted, and stemmed. The top 1% Information Gain scoring terms of the training data are used to represent pages as term-weight vectors according to the Vector Space Model, using binary weights. A linear Support Vector Machine (SVM) classifier is trained over this representation, to classify new pages as either Porn or non-Porn. The Spanish heavy filter uses the same machine learning approach, but with two additional, linguistically motivated, multi-word input features: Noun Phrases and Named Entities. Noun Phrases are recognised via part of speech tagging and regular expression matching according to a compact noun phrase grammar. Secondly, Named Entities in the training collection are recognized using a subset of the features described in the literature, focussing on persons, organizations and locations. Named Entities and Noun Phrases are taken as additional features to stoplisted, stemmed words in a SVM binary representation. We retain the 10% top IG scoring features, and learn a linear SVM classifier over the training collection, as with the light filter.

4.2.4.7 Text Filtering for French

To evaluate the ease of porting the POESIA technology to another language, the English Light Filter was augmented to handle French text. The port only relates to the Light (Statistical) Filter, as the development of a Heavy (NLP) Filter would require specific linguistic knowledge. The augmentation of the English filter involved the collecting of data (for training and testing), acquisition of a French word “stoplist” and a French stemmer. Fortunately all these resources were available from the web. The French filter was tested using approximately 5000 harmful (pornographic) and 8000 harmless (non-pornographic) pages; this data was divided evenly between the training and test sets, which resulted in a 6264 pages in the test set. The filtering performance terms of effectiveness (recall of harmful pages) is approximately 99%, whilst for over-blocking (1 – recall of harmless pages) it is approximately 6%. Note that the higher level of performance - when compared to the application of the same statistical filtering techniques for English (see section 4.3.1 below) - is due to the data used in the evaluation. Due to the limited sources of data, to acquire a reasonable sample of French pornographic pages, a higher number of pages were sampled from each site. This has increased the regularity of the pornographic material, thus leading to an improved classification performance. Whether such data adequately models the pornographic content on the WWW would require evaluation by a French user of the POESIA system, and goes beyond the project goals.

4.2.5 URL, PICS, and links based filtering

In addition to the sophisticated content-filtering techniques described in the sections above, POESIA also provides URL based filtering, PICS based filtering, and links (both static and dynamic) based filtering, all documented in D5.1 “URL, PICS and JavaScript Report” and available in specific directories under http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft.

URL filtering is a simple but useful technology. It just means that an HTTP request is filtered by the requested URL only. In POESIA some DNS domains or URL prefixes may be left unfiltered or always rejected. Also, a set of positive (= accepted) and negative (= rejected) database[4] of URLs is managed by the POESIA system. These database of URLs are actually prefixes (with the hostname reversed, so it is possible to reject any URL from the "sex.com" domain). The “URL Processing and Scoring Software Component” (D5.2) is included in the monitor. URLists are handled inside the monitor using DBM (or DB) files, in order to permit the efficient storage of millions of entries. If a URL match an URList entry, no further POESIA filtering (i.e. content analysis) is done. Currently, URLists can be loaded, dumped and managed with the POESIA monitor command. A practical major point is the collection of a significant set of negative and positive URLs, but this work is outside of the POESIA goals.

POESIA also contains a PICS filtering module, allowing or blocking access to URLs based on PICS labels describing them. PICS provides an effective and responsible way to filter Web content and helps serving diverse audiences with appropriate material. For this reason PICS has been proposed by the World Wide Web Consortium in 1995 and since then has been used and endorsed by major software industry. The PICS POESIA filter translates PICS rating labels into POESIA scores. The PICS filter scans the HTML content of the requested page for finding the PICS labels and tags in the HTML head. For each label group of a certain PICS service (identified by an URI) the filter translates the label values into POESIA scores using a XML mapping file, which associates PICS label values with POESIA domain scores. The POESIA scores obtained through this process are then passed to the monitor for them to be transferred to DM. The configuration of the module needs for each PICS service a single XML mapping file. In the POESIA system, PICS filtering is very useful since it provides an effective way for filtering web content, but unfortunately not all web pages are PICS labeled.

POESIA also contains a static links filtering module, which computes a score by getting all the static links inside an HTML page and comparing them against the database of URLs.

POESIA performs dynamic links filtering through Javascript static analysis. Since most current HTML pages contain Javascripts, it was required to do some static code analysis of these scripts (either inline, or embedded, so pre-fetched by POESIA). This code is statically analysed, because it cannot be interpreted outside of a client browser - since it depends upon user actions. By a detailed study of the actual use of Javascripts in current pages, it was assessed that usual (but quite complex) static analysis techniques such as abstract interpretation are not enough for effective analysis of Javascript. More heuristic based analysis was required to approximate the set of dynamic links reachable through the scripts. Preliminary and exploratory work on Javascript analysis was reported in D5.1 "URL, PICS and JavaScript Report"; since any complex JavaScript analysis appeared to be beyond the scope of the project, the basic implementation integrated in the POESIA system performs localised link analysis within Javascript.

[4] But building and maintaining a realistic and actual set of positive and negative URLs is outside the scope of this project, and provides significant market opportunities to European corporations.

4.2.6 Decision mechanisms

The decision mechanism component (DM) is given as input the scores associated with a given content and produces an accepting or rejecting decision.

The main focus of DM development has been put on the module architecture, and in particular on a set of tools that, independently from the algorithm used to obtain the decisions, permit the interaction with the different POESIA filters and the monitor, manipulate common operations, adapt the DM behaviour to specific user needs, e.g. for choosing i) the domains to be filtered (this is an important feature also for future improvements of the system) and ii) the filters coming into play for each domain and the MIME type. The DM component may produce a decision without having an answer from every relevant filter: for instance, a page can be rejected just because it contains strong porn images, without having completed its textual analysis. The DM component may also force a decision when a predefined time interval has passed, thus avoiding a request to wait indefinitely for a filter response. The DM algorithm may be specialized in an easy way to meet the specific needs of new domains, without modification of the general behaviour.

The DM component is configurable through a group of simple parameters, which can change the general DM behaviour by indicating, for instance, the level of filtering to be performed (e.g. for different ages), or the DM behaviour for special cases, or the time in seconds after which a forced decision has to be made. More details about DM can be found in D8.1 "DM Filtering Components Software"; DM is available at http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/Decisor.

4.3 *Evaluation of the POESIA Filtering System*

The POESIA project has successfully created an internet filtering system that combines into a single platform multiple filtering technologies which range from widely used filtering methods, e.g. positive/negative URL lists and PICS, to advanced content-based filtering technologies that address both text content and image content. Considerable effort was put in achieving a comprehensive evaluation of the POESIA system. Two different evaluations were carried out in order to assess performance, effectiveness and usability of the system developed in the project lifetime:

- a quantitative evaluation, aimed at assessing the ability of the POESIA system to correctly filter out harmful content whilst allowing harmless content to be seen;
- a qualitative evaluation, concerned with how easy the system is to install and use. Obviously when evaluating the system in these terms it has to seen in the context of a large, complex OpenSource package rather than a commercial "plug & play" piece of software.

The results of these evaluations (detailed for the different system versions respectively in D9.1 "Evaluation Report of POESIA Software Alpha Release", D9.2 "Evaluation Report of POESIA Software Beta Release" and D9.3 "Evaluation Report of POESIA Software Final Release") are summarised below.

4.3.1 Results of quantitative evaluation

The individual content filters (for the three development languages and images) were tested first in isolation, by the party responsible for the filter's development, and then the combined POESIA Filtering system was tested using a large data set of pages "spidered" from the WWW.

The quantitative results of the POESIA Filtering System are presented in terms of effectiveness and over-blocking, defined as follows:

Effectiveness= Number of harmful pages blocked/Total number of harmful pages;

Over-blocking= Number of harmless pages blocked/Total number of harmless pages.

For the combined evaluation, 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.

The following table summarises the quantitative results.

	Effectiveness			Over-blocking		
	Individual Filter	POESIA	Change	Individual Filter	POESIA	Change
English	0,952	0,969	0,017	0,034	0,019	0,015
Italian	0,953	0,966	0,013	0,020	0,050	-0,030
Spanish	0,916	0,973	0,057	0,001	0,020	-0,019
All Text Filters	0,940	0,969	0,029	0.018	0,030	-0,012
Image Only	0,910	0,933	0,023	0,200	0,041	0,159
All Filters	0,933	0,960	0,027	0,064	0,033	0,031

It is difficult to determine the causes of the changes in performance between filtering with the individual filters, and filtering with the combined POESIA system. There is the effect of the test data set (which was not the same), the interaction between the language identifier and text filters and the interaction between the text and image filters to consider.

First consider the filtering performance on pages with some textual content (note that for the POESIA system evaluation these pages might also contain image content). In general summary the blocking effectiveness for all language specific pages is approximately 97%. All three languages in the POESIA system show an improvement in effectiveness over the individual filter performance. This can, in part, be explained by the synergy between the text and image filters. Pages with a small textual content are filtered poorly by the text filters but such pages are more likely to contain other content, such as images. The average over-blocking on pages with some textual content is 3% however the performance on specific language pages is variable. The performance on Italian and Spanish pages decreases in terms of over-blocking when compared to the individual filters. However, as the English filter is used as a default by the language identifier, the decrease performance could be due to the language identification errors.

The filtering performance on image only pages, as might be expected, is less than when there is a more rich content on which to filter the page. The effectiveness is approximately 93% and over-blocking 4%. There is a considerable improvement in performance when comparing the use of the individual filter and the image filter in the POESIA system. This is due to the fact that when testing the image filter individually a single image was used whilst in the POESIA system the image filter produces its classification based on multiple images in a page, thus the increased information improves its accuracy.

Summing up, the quantitative evaluation results reported above showed that the POESIA system performs in line with the other state-of-the-art filtering systems, with effectiveness over all pages of 96% and over-blocking of 3.3%. This compares with NetProtect that provides two results depending on the decision mechanism used, either an effectiveness of 95.5% and over-blocking of 17.3%, or an effectiveness of 87.9% and over-blocking of 1.8%. Note that these figures are not directly comparable as they have different test data sets and POESIA unlike NetProtect also uses image content.

4.3.2 Results of qualitative evaluation

For the qualitative evaluation the final release of the POESIA Filtering System has been tested in a controlled environment (conducted by Telefónica I+D IT experts) and in a live environment (real End User have tested POESIA with the support of the Telefónica I+D IT experts) using different scenarios and configurations.

The system proved to be more complex to implement and run than originally envisaged, however collaboration between the developers and end-user teams enabled the end-users to refine their image of the system to see it in the more realistic context of a complex OpenSource system. Also the developers improved interaction with the software to ease its installation and use.

The installation of POESIA is a complex procedure, involving numerous packages, which requires an IT expert. However the software provides feedback to highlight the missing packages and incorrect versions of dependent software. Once installed the system is configured through the use of text files, examples of which are provided in the distribution. Again configuring the initial installation requires IT technical expertise, however simple changes to the configuration (such as changing the data used to construct a given filter's model) involve the relatively straightforward task of editing the attribute value in the specific filters configuration file.

The final version of the POESIA system has been successfully tested with various browsers under various operating systems. It was shown to be robust, and whilst there is some impact upon browsing page loading speed (especially for pages with high image content) there is an acceptable trade-off between increased filtering functionality and the slow-down introduced.

4.4 *Future improvements and extensions*

The POESIA project has achieved its ambitious central aim of producing an internet filtering system that integrates a range of filtering mechanisms, including advanced content-based filters, into a single platform. Despite this achievement, there are a number of ways in which the system could be extended or improved. Some of these extensions are ones whose usefulness could always have been anticipated, but which were beyond the scope of POESIA as a project with finite resources. Other possible improvements have come to light from the experience of creating and testing the POESIA system during the project. This is the reason why it is thus reasonable for us to list desirable improvements among the project outcomes.

There are two key topics that are crucial in setting the context for the possible future development of the POESIA system. Firstly, the OpenSource nature of the project brings the realistic likelihood that significant development effort will be applied to the system beyond the end of the project period (e.g. by interested members of the global OpenSource community). Secondly, there are architectural aspects of the POESIA system that are significant to making it feasible that further development of the system outside of the project can be successfully achieved.

In the D9.4 "Definition of Future Improvement Report", we have discussed a range of possible improvements and extensions of the POESIA system, which are listed below:

- extending filtering to other languages;
- extending filtering to other domains of harmful content, such as racial hatred, use of illegal drugs, harmful religious cults, use or creation of weapons (e.g. bomb-making);
- extending filtering to other data types (e.g. PDF and WORD files for texts as well as various video formats, such as MPEG and ASF for images);
- extending filtering to other internet channels for which it would be potentially valuable to apply filtering, e.g. news, e-mail, and chat;

- extending filtering to other data transfer media, such a CD-ROM, floppy disk, data tapes, etc.

All improvements listed above relate to extending the system's filtering capabilities in one way or another. Another set of desirable improvements comes from the experience that end-users had in working with the system. Among them it is worth mentioning: improving the ease of system installation; facilitating the reporting of bug messages; authorising local override of filtering decisions.

5 Networking and dissemination

Throughout the project lifetime, a wide variety of local, regional, national and international dissemination activities have been undertaken. These have included disseminating the project aims and outcomes, establishing emailing lists of those who wished to participate in the filter development and in receiving further information about the project. POESIA's work has been extensively reported in a number of occasions. Several presentations and communications have been undertaken about POESIA, notably by contacting several potential users (namely, persons making decisions in educational institutions) in partner countries and by attending academic and promotional conferences, exhibitions and other events. Several scientific papers, illustrating specific aspects of the project outcomes, have been presented at international conferences and workshops.

The visibility of the project was raised through a number of initiatives. At the end of the POESIA lifetime, an international workshop on "Present and Future of Open-source Content-based Web Filtering" was held in Pisa (Italy) whose intended goal was to disseminate POESIA results both in the scientific community and among the potential end-users and to benefit from personal feedback of participants external to the project. This provided a stimulating opportunity to evaluate the results achieved by the project and to compare them with respect to similar work done by other Internet content filtering teams.

POESIA's work has raised the interest of various institutions, ongoing projects and initiatives world-wide. For instance, the NCSR (Demokritos, Greece) representative of the IAP project SIFT (Konstantinos Chandrinos), invited to the final POESIA workshop, was very impressed by the live demo of the POESIA filtering system ("one of the best filters") and offered its integration into the ICRAplus architecture where several filters can be exploited in parallel. Recently, the POESIA filtering system attracted the attention of a big company in Italy offering support to primary and secondary schools in the North of Italy for what concerns educational networks, internet access, opensource software installation; preliminary contacts are now being conducted.

POESIA has also been actively promoted at trans-European level to end-users. Press releases were mailed to more than 800 european contacts and the results have been published on key Web sites and via electronic newsletters including eLearning Europa, European Schoolnet, Safeborders and EDEN (European Distance Education Network). The outcomes of POESIA were also presented at the Safer Internet forum in Poland as part of the Safeborders project and in Bucharest, Romania at the E-COMM LIN conference, meeting the mission to embrace the enlargement countries. Leaflets were distributed at many European information technology and education workshops, seminars, exhibitions and conferences for schools, such as those organised by RIAC (Comenius school network), MET (Mobile European Teacher), BETT (British Education Technology Show), ILT (Institute for Learning Technology) and the EDEN annual conference. A teacher workshop was held in Liverpool as part of Safer Internet Day (February 6th). A POESIA presentation was also made at the Council of Europe January 2004 meeting for Educational NGOs in Strasbourg, France.

5.1 *The POESIA Websites for Dissemination and OpenSource Development*

Since the very beginning of the project, the POESIA Consortium set up a dedicated dissemination website (see Figure 2 below).

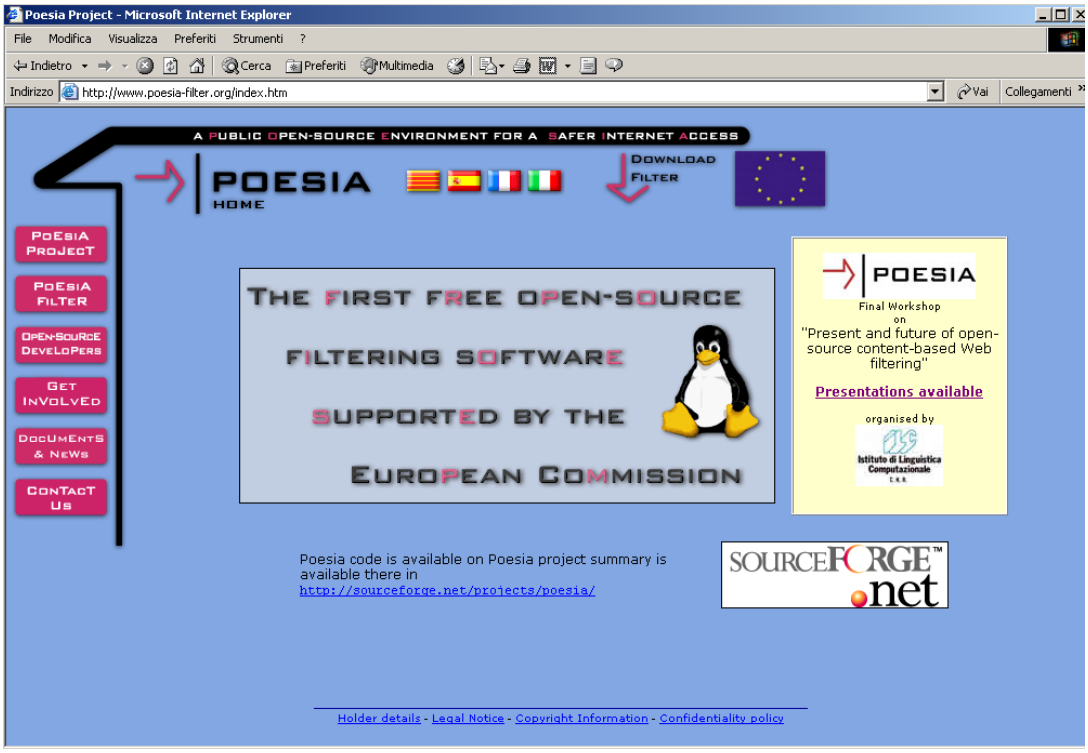

Figure 2: The home page of the POESIA dissemination web site

This website – accessible on http://www.poesia-filter.org/index.htm - has been maintained and constantly updated during all the course of the project. All deliverables – both public and restricted ones (the latter accessible with password) – have been available for download as PDF documents. This site will be maintained at least until the end of 2005.

This site also provides a link through to the POESIA opensource development site at SourceForge (http://sourceforge.net/projects/poesia/), from which the POESIA source code can be downloaded. SourceForge is the world's largest Open Source software development website, hosting development websites for a large number of OpenSource projects. Each project site makes the project applications freely available to potential users, but also they provide a central focus for development work. In particular, open source developers can download and modify the current version of the source code, and submit their improvements for inclusion in the system. It is through this cycle of modification and inclusion that opensource development proceeds, to produce rapid system evolution. Sites also host project specific mailing lists, to allow effective communication amongst developers, and between developers and users. With sufficient interest from enthusiastic opensource developers, there is no reason why the POESIA development website at SourceForge should not continue indefinitely.

6 Conclusions

The POESIA project has achieved its ambitious central aim of producing an open-source internet filtering system that integrates a range of filtering mechanisms, including advanced content-based filters, into a single platform. The POESIA system aims to be the standard Open Source solution for

Internet content filtering of harmful content to minors. The fact that this large and complex software system has been delivered in functioning order within a narrow two-year project period is a testament to the hard work and commitment of the project participants.

The current POESIA filtering system covers English, Italian, Spanish and French (the latter only to a limited extent). In terms of filtered domains, pornography is the main issue, extended to gross language (in text) and to the identification of material having violent/hate associations (through the recognition in images of symbols that are logos for relevant groups). The system was extensively tested against various sets of URLs, both in controlled and live environments, using different scenarios and configurations, and always showed a performance in line with other state-of-the-art filtering systems.

Due to the OpenSource nature of the project and to the architectural characteristics of the system, POESIA filtering software is set to take advantage of future development and improvements. There are many conceivable extensions of functionality and operation that would constitute valuable additions to the system which have been delivered among the project outcomes. We can thus conclude that the POESIA system has a promising future, in which the system is likely to undergo continued development, so as to maintain its status as a state-of-the-art filtering system deployed in educational establishments to protect children from exposure to harmful material.

Activities in POESIA have resulted in the submission of other project applications. Members of the project team (FCR, Liverpool Hope) submitted a successful project application for an Internet Awareness project Safeborders. Working with POESIA provided a good understanding of the issues concerned. Members of the POESIA consortium have also submitted an application for the development of a teacher-training course on Internet Safety to the Socrates-Comenius action. Other funding like Framework 6 (IST) is also being considered.

7 Appendix 1: Deliverable List

Deliverable No.	Deliverable Title	Delivered	Status	Date Due	Availability
D 1.1.a	Progress Report 1	yes	Restricted	t6	http://www.poesia-filter.org/pdf/Deliverable_1_1_a.pdf
D 1.1.b	Progress Report 2	yes	Restricted	t12	http://www.poesia-filter.org/pdf/Deliverable_1_1_b.pdf
D 1.1.c	Progress Report 3	yes	Restricted	t18	http://www.poesia-filter.org/pdf/Deliverable_1_1_c.pdf
D 1.2.a	Management Report 1	yes	Restricted	t3	http://www.poesia-filter.org/pdf/Deliverable_1_2_a.pdf
D 1.2.b	Management Report 2	yes	Restricted	t9	http://www.poesia-filter.org/pdf/Deliverable_1_2_b.pdf
D 1.2.c	Management Report 3	yes	Restricted	t15	http://www.poesia-filter.org/pdf/Deliverable_1_2_c.pdf
D 1.2.d	Management Report 4	yes	Restricted	t21	http://www.poesia-filter.org/pdf/Deliverable_1_2_d.pdf
D 1.3	Annual Report	yes	Public	t12	http://www.poesia-filter.org/pdf/Deliverable_1_3.pdf
D 1.4	Final Report	yes	Restricted	t24	http://www.poesia-filter.org/pdf/Deliverable_1_4.pdf
D 2.1	End-user Requirements Report	yes	Public	t6, t12	http://www.poesia-filter.org/pdf/Deliverable_2_1.pdf
D 2.2	Test Case Files	yes	Restricted	t6, t12	http://www.poesia-filter.org/pdf/D22.zip
D 3.1	Software Architecture Definition Document	yes	Public	t6, t9	http://www.poesia-filter.org/pdf/Deliverable_3_1.pdf
D 4.1	Alpha Release of POESIA Software	yes	Public	t12	http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/
D 4.2	Beta Release of POESIA Software	yes	Public	t18	http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/
D 4.3	Final Release of POESIA Software	yes	Public	t24	http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/
D 5.1	URL, PICS and JavaScript Report	yes	Public	t9, t18	http://www.poesia-filter.org/pdf/Deliverable_5_1.pdf
D 5.2	URL Processing and Scoring Software Component	yes	Public	t15	http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/UrlFilter/
D 5.3	JavaScript Analysis Software Component	yes	Public	t18, t24	http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/TextFilter/English/src/org/poesia/filters/javascript

Deliverable No.	Deliverable Title	Delivered	Status	Date Due	Availability
D 6.1	Image Subsets	yes	Restricted	t12	A CD with image subsets is available and was sent by s-mail to the Commission
D 6.2	Skin Feature Extraction and Detection Component	yes	Public	t15, t21	http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/ImageFilter/
D 7.1	Domain Corpora for Each Language	yes	Restricted	t12, t18	A CD with collected domain specific corpora is available and was sent by s-mail to the Commission
D 7.2	Lexical Resources and Tools for Each Language	yes	Public	t12	http://www.poesia-filter.org/pdf/Deliverable_7_2.pdf
D 7.3	Crude Text Filter Software	yes	Public	t15, t21	http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/TextFilter/English/ http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/TextFilter/Italian/ http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/TextFilter/Spanish/
D 7.4	Elaborate Text Filter Software	yes	Public	t18, t24	http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/TextFilter/English/ http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/TextFilter/Italian/ http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/TextFilter/Spanish/
D 7.5	Partial Porting to French Language	yes	Public	t24	http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/TextFilter/English/src/org/poesia/utils/french http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/TextFilter/English/src/org/poesia/filters/french
D 8.1	DM Filtering Components Software	yes	Public	t12, t21	http://www.poesia-filter.org/pdf/POESIA_IAP-2117-27572_D8.1.pdf http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/Decisor
D 8.2	Software Tools for the Configuration and Administration of these FDMM	yes	Public	t18, t24	http://cvs.sourceforge.net/viewcvs.py/poesia/PoesiaSoft/Decisor
D 9.1	Evaluation Report of POESIA Software Alpha Release	yes	Restricted	t15	http://www.poesia-filter.org/pdf/Deliverable_9_1.pdf
D 9.2	Evaluation Report of POESIA Software Beta Release	yes	Restricted	t21	http://www.poesia-filter.org/pdf/Deliverable_9_2.pdf
D 9.3	Evaluation Report of POESIA Software Final Release	yes	Restricted	t24	http://www.poesia-filter.org/pdf/Deliverable_9_3.pdf passwd: homerus

Deliverable No.	Deliverable Title	Delivered	Status	Date Due	Availability
D 9.4	Definition of Future Improvement Report	yes	Public	t24	http://www.poesia-filter.org/pdf/Deliverable_9_4.pdf
D 9.5	Dissemination Report and WWW Server	yes	Public	t9	http://www.poesia-filter.org/pdf/Deliverable_9_5.pdf
D 9.6	Final CD-ROM of POESIA Results	yes	Public	t24	A CD with the final version of the POESIA system is available and was sent by s-mail to the Commission