

Encyclopedia of Internet Technologies and Applications

Mario Freire
University of Beira Interior, Portugal

Manuela Pereira
University of Beira Interior, Portugal

Information Science
REFERENCE

INFORMATION SCIENCE REFERENCE

Hershey • New York

Acquisitions Editor: Kristin Klinger
Development Editor: Kristin Roth
Senior Managing Editor: Jennifer Neidig
Managing Editor: Sara Reed
Copy Editor: Larissa Vinci and Mike Goldberg
Typesetter: Amanda Appicello and Jeffrey Ash
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: <http://www.igi-global.com/reference>

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: <http://www.eurospanonline.com>

Copyright © 2008 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher.

Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Encyclopedia of Internet technologies and applications / Mario Freire and Manuela Pereira, editors.
p. cm.

Summary: "This book is the single source for information on the world's greatest network, and provides a wealth of information for the average Internet consumer, as well as for experts in the field of networking and Internet technologies. It provides the most thorough examination of Internet technologies and applications for researchers in a variety of related fields"--Provided by publisher.

Includes bibliographical references and index.

ISBN 978-1-59140-993-9 (hardcover) -- ISBN 978-1-59140-994-6 (ebook)

I. Internet--Encyclopedias. I. Freire, Mário Marques, 1969- II. Pereira, Manuela.

TK5105.875.I57E476 2007

004.67'803--dc22

2007024949

British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this encyclopedia set is original material. The views expressed in this encyclopedia set are those of the authors, but not necessarily of the publisher.

Advanced Techniques for Web Content Filtering

Elisa Bertino

Purdue University, USA

Elena Ferrari

Università degli Studi dell'Insubria, Italy

Andrea Perego

Università degli Studi dell'Insubria, Italy

Gian Piero Zarri

Université Paris IV, France

INTRODUCTION

Web content filtering concerns the enforcement of mechanisms able to inform users about the content and characteristics of the Web resources they access. The former application of such techniques focused on blocking access to resources was considered inappropriate to given users (e.g., children). Nonetheless, its possible applications are not limited to such issue since it can be effectively used in order to notify users about the “quality” of a resource (i.e., the fact of satisfying given requirements), which may concern the authoritative-ness and/or authenticity of the provided information, the treatment of personal data, and so on.

In this chapter, besides discussing the current strategies for Web content filtering, and outlining their advantages and drawbacks, we present an approach, formerly developed in the framework of the EU project EUFORBIA, which, besides addressing the main drawbacks of the existing systems, can be applied for purposes comprising both users’ protection and quality assurance. The main features of such an approach are the support for multiple metadata vocabularies for the rating and filtering of Web resources, and the possibility of specifying policies, which allow the system to decide whether a resource is appropriate or not for a given user based on his or her preferences and characteristics.

BACKGROUND

Classification of Web resources based on their content emerged as a need as soon as online information exponentially grew in quantity and heterogeneity during the 1990s. Search engines were designed in order to simplify to users the task of finding what they were looking for. Nonetheless, another issue to be addressed was (and still is) making users able to realize whether a Web resource satisfies a given set of requirements concerning its content and/or characteristics. An example of such requirements concerns protecting children from possible harmful contents. Search engines are not suitable for this purpose since they cannot ensure users of having correctly rated a resource (i.e., they provide only a probabilistic classification of Web resources) and, consequently, they may consider as appropriate resources that are, by contrast, inappropriate. For this reason, this issue has been addressed by building tools able to filter Web content, which adopted two main strategies. The former is based on the manual classification of resources in order to build white and/or black lists, which are then used in order to verify whether the user is requesting access to an appropriate or inappropriate resource. The latter follows the PICS approach (PICS, 2005; Resnick, & Miller, 1996), where resources are rated by associating with them labels containing a formal description of their content/characteristics. Based on such labels and on the user profile, the filtering tool is

then able to decide whether a requested resource should be considered as appropriate or not.

Both such approaches have a main drawback in that they focus mainly on filtering efficiency. As a result, they adopt strategies that enforce a very restrictive access to online resources, with the consequence that users' navigation is limited to a very narrow subset of the Web. The label-based approach has been designed to solve such problem. Nonetheless, the supported metadata vocabularies are semantically poor, and they are not suitable for supporting policies more flexible and expressive than those used in the list-based approaches. This is why, so far, Web content filtering has been applied mainly for minors' protection purposes, although it may have other relevant applications that can help in improving users' confidentiality in using the information and services available in the Web. In particular, in recent years, it emerged the need of designing technologies able to ensure the "quality" of the Web. The P3P W3C standard (Cranor, Langheinrich, Marchiori, Presler-Marshall, & Reagle, 2002) has been one of the first outcomes of such research trend; thanks to this technology, users can be aware of how their personal data will be used by a Web service, and thus they can decide whether to use it or not. Nonetheless, besides privacy protection, the quality of Web resources can be ensured with respect to a variety of requirements, which may help users in being more confident with the Web. Suppose, for instance, that a user accesses a medical Web site, providing information about diseases, their symptoms, and the corresponding treatments; currently, such user cannot be sure if such information is reliable and accurate, and thus he or she must avoid considering it for addressing his or her needs. The availability of more sophisticated filtering systems may also help in addressing such issues.

THE EUFORBIA APPROACH TO WEB CONTENT FILTERING

In this section, we illustrate a filtering framework, formerly developed in the context of the EU project EUFORBIA (EUFORBIA, 2004), which addresses the main drawbacks of existing filtering systems; on one hand, by supporting both the list- and metadata-based approaches, and, on the other hand, by allowing the specification of policies taking into account the characteristics of both users and resources. The main com-

ponents of our framework are a cross-domain ontology, allowing an accurate description of Web resources, a new format for labels, which describe both the content and the structure of Web sites, and two integrated systems, specifically designed in order to address the needs of both home and institutional users.

THE EUFORBIA ONTOLOGY AND EUFORBIA LABELS

To make it possible to apply Web content filtering for purposes not limited to users' protection, it is necessary to have the proper tools for describing resources from different points of view. This can be obtained by supporting multiple domain-dependent metadata vocabularies, focusing on a specific usage of Web resources.

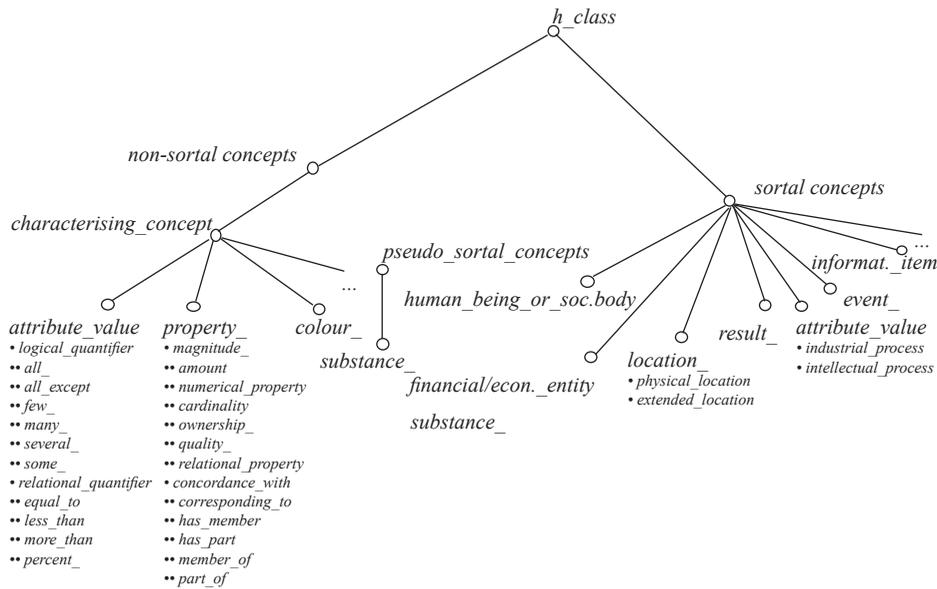
Besides supporting such features, we designed a cross-domain ontology, which can be extended with concepts pertaining to new content domains. More precisely, the EUFORBIA ontology has been obtained by extending the general class hierarchy (HClass) of NKRL (Zarri, 2003) with concepts pertaining to the domain of users' protection (e.g., pornography). Figure 1 depicts a portion of the upper level of the HClass hierarchy.

The main difference between NKRL and the usual ontology paradigm is the support for an ontology of events in addition to the traditional ontology of concepts. In an ontology of events, nodes represent multiadic structures called *templates* obtained by combining quadruples consisting of the template name, a predicate, and the arguments of the predicate, which consist of a structured set of concepts.

The NKRL ontology of events is a hierarchy of templates (HTemp), which correspond to classes of elementary events such as move a physical object, be present in a place, produce a service, send/receive a message, build up an Internet site, and so forth. Currently, HTemp provides about 180 templates, which can be instantiated in order to represent an actual event (e.g., "Mr. Smith has created a new Web site having the URL ..."). Templates represent, in short, all sort of *dynamic relationships* between basic concepts, and are then of vital importance for the correct rendering of *narratives*—in the widest meaning of this term, including also, the semantic content of Web sites.

Note that there is no real contradiction between NKRL and the ontological languages such as OWL

Figure 1. An abridged representation of the HClass hierarchy



Box 1.

```

c73) (COORD c76 c77)
c76) OWN SUBJ SNOOP_DOGG_INTERNET_SITE (www.snoop-dogg.com)
      OBJ property_
      TOPIC (SPECIF dedicated_to SNOOP_DOGG)
// instance of: Own:CompoundProperty
c77) BEHAVE SUBJ SNOOP_DOGG (www.snoop-dogg.com)
      MODAL rap_star
// instance of: Behave:Role
    
```

(McGuinness & Harmelen, 2004). This is due mainly to the fact that the ontology of events (HTemp) of NKRL works strictly in conjunction with an NKRL standard ontology of concepts (HClass)—as already stated, the arguments of the predicate in the templates are built up from concepts or combination of concepts. Moreover, translators from/to XML (Bray, Paoli, Sperberg-McQueen, Maler, & Yergeau, 2006) and RDF(s) (Brickley & Guha, 2004; Klyne & Carroll, 2004) exist for NKRL. Extensions to RDF have been developed in order to preserve all the NKRL power when NKRL structures are converted into RDF (e.g., by securing full compatibility between RDF containers and NKRL expansion operators).

NKRL and the EUFORBIA ontology are the tools on which EUFORBIA labels are based. More precisely,

labels are expressed in NKRL, using the set of concepts provided by the EUFORBIA ontology. A EUFORBIA label consists of three main sections: the first (aims) describes the Web site purposes, the second (properties) provides information about some relevant characteristics of a Web site, whereas the third (sub-sites) illustrates the structure of the Web site.

A very simple example of the aim section of a EUFORBIA label is reproduced in Box 1. Such set of assertions can be translated as follows: “The site is devoted to an individual named Snoop Dogg; this individual is a rap star.” Statement c73 specifies that this section of the labels consists of two “predicative occurrences,” c76 and c77, that must take place together (COORD = coordination). Statements c76 and c77 are instances of two different NKRL templates, respectively, Own:

Box 2.

```

Name: Behave:Role
Father: Behave
Position: 1.1
NL description: "A Human Being or a Social Body Acts in a Particular Role"

BEHAVE SUBJ var1: [(var2)]
*OBJ
[SOURCE var3: [(var4)]]
[BENF var5: [(var6)]]
MODAL var7
[TOPIC var8]
[CONTEXT var9]
[ modulators ]

var1 = <human_being_or_social_body>
var2 = <geographical_location> | <premise_>
var3 = <human_being_or_social_body>
var5 = <human_being_or_social_body>
var7 = <role_>
var8 = <entity_>
var9 = <situation_> | <symbolic_label>
var4, var6 = <geographical_location>
    
```

Compound Property and Behave:Role. A representation of the latter is depicted in Box 2.

As can be seen, the arguments of the predicate are expressed through variables (**vari**) and constraints on the variables, like *human_being_or_social_body* for **var1**.

THE MFilter PROTOTYPE

The first EUFORBIA prototype, referred to as MFilter (Bertino, Ferrari, & Perego, 2003, 2006), is based on a filtering model where policies are specified on either the identity or characteristics of users and resources. Resources’ characteristics are described by labels, which may use different metadata vocabularies (thus, also the EUFORBIA ontology). Metadata vocabularies are also used for describing user profiles. In such a case, the adopted vocabulary depends on the context (e.g., a user profile in a school context will be different with the profiles adopted in companies). An example of a hierarchy of user profiles in a school context is depicted in Figure 2.

This is a relevant improvement with respect to the available filtering systems where users are grouped into one or more static classes, taking into account only a

given parameter (e.g., age). By contrast, in our model the classes of resources and users to which a policy applies are determined dynamically in policies by expressing constraints on their properties. Thus, we can specify a policy stating that the sexual content is inappropriate for students of less than 15 years old.

Two other relevant features of our model are the support for both negative and positive policies, and the support for a policy propagation mechanism. Thanks to the former feature, we can specify policies stating not only which resources are inappropriate (negative sign), but also those that are appropriate (positive sign) for a given user. The latter feature is related to the fact that metadata vocabularies may be hierarchically organized (as the EUFORBIA ontology, and the user profiles in Figure 2). Thus, we can exploit such hierarchy according to the principle that a policy applying to a given, say, user profile, is inherited by all its children.

Although such features improve the expressivity of our model, they may cause conflicts between policies (e.g., two policies applying to the same user and the same resource, but with different sign). For this reason, our model supports also a conflict resolution mechanism, based on the principle of stronger policy, according to which the prevailing policy is the more specific one wrt the hierarchy. In case the policies are

Figure 2. An example of user profile hierarchy

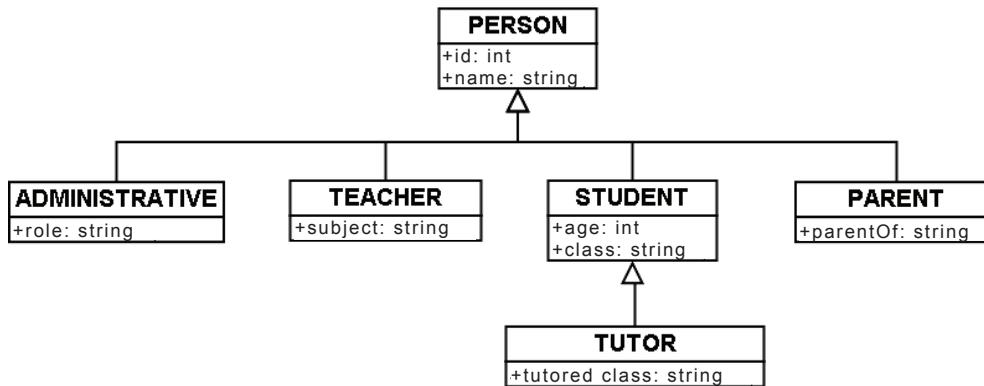
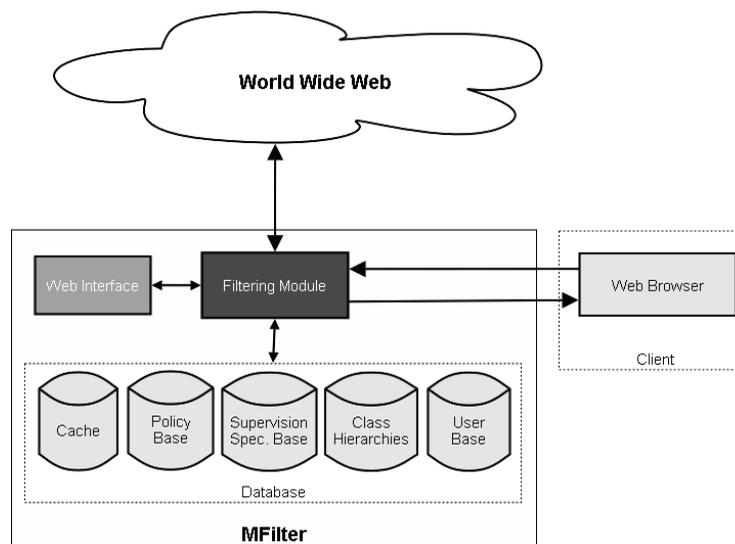


Figure 3. MFilter architecture



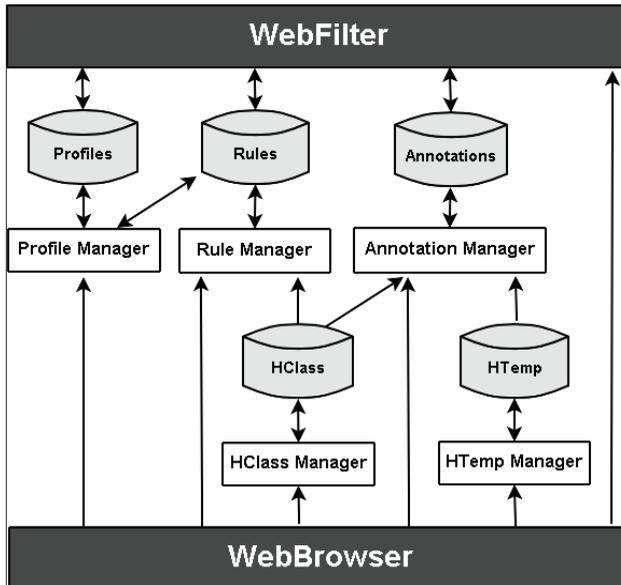
incomparable with respect to the hierarchy, the prevailing policy is the one with the stronger sign, which may be either the positive or the negative one, depending on the specific context.

Finally, our model provides support for concurrent policy specification (i.e., it allows different classes of supervisors, with possibly different authority levels) to specify policies applying to the users of whom they are responsible. This feature is useful since it may often be the case that the responsibility of deciding what is appropriate or not for a user is shared among different persons (e.g., in a school context, teachers and parents). In case two supervisors specify conflicting policies, their authority level is taken into account in order to determine the prevailing policy.

The architecture of MFilter, depicted in Figure 3, consists of three main components: a database, storing all the data needed by the system, a Web interface, for managing the system, and the filtering module, which is in charge of processing all the access requests and verifying whether they are appropriate or not based on the policies stored in the database.

Since the filtering module must evaluate a high number of access requests and, for each of them, identify the set of policies applying to the requesting user and the requested resource, this may affect both the efficiency and effectiveness of filtering. It is in fact fundamental that the response delay is not perceived by users. For this purpose, pre-computational strategies and caching

Figure 4. Web browser module of the NKRL-based EUFORBIA prototype



mechanisms are enforced by the filtering module. More precisely, the filtering module stores the results of the evaluations already carried out in the MFilter database (see below) so that they can be re-used whenever the corresponding access requests are submitted again, thus avoiding to perform the filtering procedure.

The main components of the MFilter database schema are depicted in Figure 3. The rating system, user, and policy bases store all the corresponding data directly inserted by the system administrator (SA). Currently, MFilter supports all the PICS-based rating systems and the EUFORBIA ontology. The resource base stores the URIs and metadata concerning the already requested resources as previously mentioned. Yet, since Web page content is dynamic and may change with a frequency depending on the Web page type, all the metadata concerning online resources are associated with an expiring time set by default or possibly specified by the SA. Moreover, the cache component stores all the pre-computed data regarding the policies, but also all the access requests already submitted along with the identifier of the users and the corresponding (negative or positive) responses. All this information is accessed and managed by the SA through the Web interface.

Finally, the MFilter Web interface allows the SA to insert, edit, and/or delete data stored in the database. It is structured into three main sections: the administration interface, for the management of the filtering system; the supervision interface, for validating the policies specified by the SA; the user interface, the aims of which are (a) to perform user authentication, (b) to supply some general information about the prototype and its use, and (c) to provide a tool for editing the user's password.

THE WebBrowser AND THE INFERENCE RULES

The architecture of the second EUFORBIA prototype, depicted in Figure 4, consists of two main components, namely the *WebBrowser* and *WebFilter*. Such software, installed on a Proxy server or on the user's machine, performs the basic filtering actions of this prototype.

By using the *ProfileManager* module, the user can specify the categories of resources that he or she considers as appropriate and inappropriate. Such "user profile" is stored in the "profile" repository. The "rules" repository stores the existing filtering rules expressed in NKRL. Whenever a EUFORBIA label is retrieved, it is unified with the left-hand side of the rules applying to the requesting user. If the label can be unified with the rules, the system returns a response corresponding to "resource *r* is inappropriate for user *u*."

The information provided by EUFORBIA labels allows us to enforce filtering with a granularity finer than the whole Web site. In other words, we may have different evaluation results depending on the possible presence of Web pages or Web site sub-sections different with respect to their content.

In Box 3, the NL transcription of some simple filtering rules is provided. The actual code of RULE2 is reported in Box 4.

Note that in spite of the (apparently) different knowledge representation formalisms used, the concrete deductive power of these rules does not seem to be very different from that of other rule systems proposed recently in a Semantic Web context (Shadbolt, Berners-Lee, & Hall, 2006; SW, 2006) as SWRL (Horrocks et al., 2004).

The particular inference engine used in the NKRL-based prototype of EUFORBIA is a simplified version of the standard NKRL inference engine, based on a

A

Box 3.

Only one condition ($IF\ cond1 \Rightarrow deny\ access$)
 RULE1: *IF* a given site/site section praises topics like racism, violence, sexuality, etc.
THEN the site/site section must be refused
 (see, e.g., the sub-section lyrics.shtml of www.snoop-dogg.com site that advocates sexuality, violence and drug legalization).

Two conditions ($IF\ cond1 \wedge\ cond2 \Rightarrow deny\ access$)
 RULE1: *IF* a given site/site section is devoted to spreading information over the Internet
 AND this information concerns sexuality, violence, racism, etc.
THEN the site/site section must be refused
 (see, e.g., the site www.textgais.com, devoted to the diffusion of short stories about homosexuality).

Three conditions ($IF\ cond1 \wedge\ cond2 \wedge\ cond3 \Rightarrow deny\ access$)
 RULE1: *IF* a given site does not contain explicit nudity
 AND this site includes sections
 AND some of these show photos of women/teenagers dressed in bras or underwears
THEN the site (or site sections, according to the chosen strategy) section must be refused
 (see, e.g., sections “mixed-sheer-bras” and “mixed-teen-bras” of softcoregp.com/bracity).

Box 4.

```

RULE2:COND1) OWN SUBJ var1
    OBJ property_
    TOPIC (SPECIF dedicated_to (SPECIF internet_publishing var2))
RULE2:COND2) OWN SUBJ var2
    OBJ property_
    TOPIC var3

var1 = site_ | internet_site_section
var2 = information_item
var3 = sexuality_ | violence_ | racism_

```

classical backward chaining approach with chronological backtracking. The differences with respect to other well-known examples of use of this approach (Mycin, PROLOG, etc.) are mainly linked with the complexity of the NKRL data structures. This complexity implies the execution of difficult operations of reconstruction of the program environment whenever, after a deadlock, it is necessary to return to the previous choice point to try a new way of pursuing with the processing.

FUTURE TRENDS

We are currently investigating how our approach can be enforced in the framework of the semantic Web and in the W3C Web services architecture (Booth et al., 2004), where the possible applications of Web content filtering will play a relevant role in customizing the Web to the

needs and preferences of end users. A first step toward this use of Web content filtering has been established in the framework of the EU project QUATRO (QUATRO, 2006), where a standard RDF format for content labels and trust marks has been defined. Besides enhancing the metadata-based approach to Web content rating and filtering, QUATRO addresses one of the major issues in such field, namely, verifying the accuracy and trustworthiness of content labels. This is obtained by applying flexible strategies for labels’ authentication and validation, which can be tailored to the labeling policy of the organization which released the labels. Moreover, an integrated system architecture has been designed and implemented, consisting of a set of Web services, in charge of labels’ retrieval and validation, and two front-end tools, namely, ViQ and LADI, which notify to end users the evaluation results concerning labels’ content and validity (Karkaletsis et al., 2006).

A relevant outcome of the project has been the establishment of a W3C incubator activity (Web Content Label Incubator Group, 2006), aiming at proposing a W3C standard for Web content labels, starting from the QUATRO experience.

CONCLUSION

Web content filtering aims at analyzing the content/ characteristics of Web resource in order to verify whether they satisfy the preferences and requirements expressed by end users. In this chapter, we discussed the possible applications of Web content filtering, and how it is currently enforced, outlining the advantages and drawbacks of both the list- and metadata-based approaches.

As an example of how Web content filtering can be enhanced, we illustrated the EUFORBIA filtering approach, the main features of which are the support for both list- and metadata-based approaches, the support for multiple metadata vocabularies for the rating and filtering of Web resources, and the possibility of specifying policies, which allow the system to decide whether a resource is appropriate or not for a given user, based on his or her preferences and characteristics.

Finally, we briefly presented a novel approach developed in the context of the QUATRO project, where Web content filtering is enforced by using a metadata-based approach based on Semantic Web technologies. Such an approach indicates the future trends of Web content filtering, namely, its integration into the Semantic Web and Web services frameworks, where it will play a relevant role in customizing the Web to the needs and preferences of end users.

REFERENCES

- Bertino, E., Ferrari, E., & Perego, A. (2003, November). Content-based filtering of Web documents: The MaX system and the EUFORBIA project. *International Journal of Information Security*, 2(1), 45-58.
- Bertino, E., Ferrari, E., & Perego, A. (2006). Web content filtering. In E. Ferrari & B. Thuraisingham (Eds.), *Web and information security* (pp. 112-132). Hershey, PA: Idea Group Publishing.
- Booth, D., Haas, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C., et al. (2004, February). *Web services architecture* (W3C Working Group Note). Retrieved from <http://www.w3.org/TR/ws-arch>
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., & Yergeau, F. (2006, August). *Extensible markup language (XML) 1.0. Fourth Edition* (W3C Recommendation). Retrieved from <http://www.w3.org/TR/xml>
- Brickley, D., & Guha, R. (2004, February). *RDF vocabulary description language 1.0: RDF Schema* (W3C Recommendation). Retrieved from <http://www.w3.org/TR/rdf-schema>
- Cranor, L., Langheinrich, M., Marchiori, M., Presler-Marshall, M., & Reagle, J. (2002, April). *The Platform for Privacy Preferences 1.0 (P3P1.0) specification* (W3C Recommendation). Retrieved from <http://www.w3.org/TR/P3P>
- The EUFORBIA project* (Project Web Site). (2004, September). (<http://semioweb.msh-paris.fr/euforbias>)
- Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosz, B., & Dean, M. (2004, May). *SWRL: A Semantic Web rule language combining OWL and RuleML* (W3C Member Submission). World Wide Web Consortium. Retrieved from <http://www.w3.org/Submission/SWRL>)
- Karkaletsis, V., Perego, A., Archer, P., Stamatakis, K., Nasikas, P., & Rose, D. (2006). Quality labeling of Web content: The QUATRO approach. In *Proceedings of the WWW'06 Workshop on Models of Trust for the Web (MTW 2006)*. Retrieved from http://www.l3s.de/~olmedilla/events/MTW06_papers/paper11.pdf
- Klyne, G., & Carroll, J. J. (2004, February). *Resource description framework (RDF): Concepts and abstract syntax* (W3C Recommendation). Retrieved from <http://www.w3.org/TR/rdf-concepts>
- McGuinness, D. L., & Harmelen, F. van (2004, February). *OWL Web ontology language overview* (W3C Recommendation). Retrieved from <http://www.w3.org/TR/owl-features>
- Platform for Internet Content Selection (PICS)* (W3C Initiative Web Site). (2005, July). (<http://www.w3.org/PICS>)
- The QUATRO project* (Project Web Site). (2006, December). (<http://www.quatro-project.org>)

Resnick, P., & Miller, J. (1996, October). PICS: Internet access controls without censorship. *Communications of the ACM*, 39(10), 87-93.

Semantic Web (W3C Activity Web Site). (2006, October). (<http://www.w3.org/2001/sw>)

Shadbolt, N., Berners-Lee, T., & Hall, W. (2006). The Semantic Web revisited. *IEEE Intelligent Systems*, 21(3), 96-101.

Web Content Label Incubator Group (W3C Incubator Activity Web Site). (2006, July) (<http://www.w3.org/2005/Incubator/wcl>)

Zarri, G. P. (2003). A conceptual model for representing narratives. In R. Jain, A. Abraham, C. Faucher, & B. van der Zwaag (Eds.), *Innovations in knowledge engineering*. Adelaide: Advanced Knowledge International.

KEY TERMS

List-Based Web Rating: A classification of Web resource with respect to their content and/or characteristics, into two distinct groups, corresponding to resources considered as appropriate (white lists) or inappropriate (black lists) for a given set of end users.

Metadata Vocabulary: A formal definition of a set of descriptors to be used for denoting the characteristics of resources (e.g., an ontology is a metadata vocabulary). Usually, metadata vocabularies are domain-specific.

Metadata-Based Web Rating: A manual or semi-automatic description of Web resources with respect to their content and/or characteristics, based on a set of descriptors defined by a *metadata vocabulary*.

Web Content Filtering: The analysis of the content and/or characteristics of Web resource with respect to the preferences expressed by end users. Such analysis is performed based on the resource classification carried out either by a *list-* or *metadata-based Web rating* approach.

Web Content Label: A formal description of the content and/or characteristics of Web resources, by using descriptors defined in one or more *metadata vocabularies*. Currently, content labels are encoded by using two W3C standards, namely, PICS and RDF/OWL.

Web Content Rating: The classification of Web resources with respect to their content and/or characteristics. Resource classification is performed by using two orthogonal strategies, namely, *list-based Web rating* and *metadata-based Web rating*.

Web Trust Mark: A third-party certification that a resource satisfies a given set of requirements (e.g., the VeriSign seal is a Web trust mark). A trust mark may be just a graphical symbol attached to the resource or even a *Web content label*.