



IRM PRESS

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-<http://www.irm-press.com>

ITB11702

This chapter appears in the book, *Web and Information Security*
edited by Elena Ferrari and Bhavani Thuraisingham © 2006, Idea Group Inc.

Chapter VI

Web Content Filtering

Elisa Bertino, Purdue University, USA

Elena Ferrari, University of Insubria at Como, Italy

Andrea Perego, University of Milan, Italy

Abstract

The need to filter online information in order to protect users from possible harmful content can be considered as one of the most compelling social issues derived from the transformation of the Web into a public information space. Despite that Web rating and filtering systems have been developed and made publicly available quite early, no effective approach has been established so far, due to the inadequacy of the proposed solutions. Web filtering is then a challenging research area, needing the definition and enforcement of new strategies, considering both the current limitations and the future developments of Web technologies—in particular, the upcoming Semantic Web. In this chapter, we provide an overview of how Web filtering issues have been addressed by the available systems, bringing in relief both their advantages and shortcomings, and outlining future trends. As an example of how a more accurate and flexible filtering can be enforced, we devote the second part of this chapter to describing a multi-strategy approach, of which the main characteristics are the integration of both list- and metadata-based

techniques and the adoption of sophisticated metadata schemes (e.g., conceptual hierarchies and ontologies) for describing both users' characteristics and Web pages content.

Introduction

In its general meaning, information filtering concerns processing a given amount of data in order to return only those satisfying given parameters. Although this notion precedes the birth of the Internet, the success and spread of Internet-based services, such as e-mail and the Web, resulted in the need of regulating and controlling the network traffic and preventing the access, transmission, and delivery of undesirable information.

Currently, information filtering is applied to several levels and services of the TCP/IP architecture. Two typical examples are spam and firewall filtering. The adopted strategies are various, and they grant, in most cases, an efficient and effective service. Yet, the filtering of online multimedia data (text, images, video, and audio) is still a challenging issue when the evaluation of their semantic meaning is required in order to verify whether they satisfy given requirements. The reason is that the available techniques do not allow an accurate and precise representation of multimedia content. For services like search engines, this results in a great amount of useless information returned as a result of a query. The problem is much more serious when we need to prevent users from retrieving resources with given content (e.g., because a user does not have the rights to access it or because the content is inappropriate for the requesting user). In such a case, filtering must rely on a thorough resource description in order to evaluate it correctly.

The development of the Semantic Web, along with the adoption of standards such as MPEG-7 (TCSVT, 2001) and MPEG-21 (Burnett et al., 2003), may seemingly overcome these problems in the future. Nonetheless, currently online information is unstructured or, in the best case, semi-structured, and this is not supposed to change in the next few years. Thus, we need to investigate how and to what extent the available techniques can be improved to allow an effective and accurate filtering of multimedia data.

In this chapter, we focus on filtering applied to Web resources in order to avoid possibly harmful content accessed by users. In the literature, this is usually

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/web-content-filtering/31085

Related Content

Balanced Approach for Hiding Sensitive Association Rules in Data Sharing Environment

Janakiramaiah Bonamand Ramamohan Reddy (2014). *International Journal of Information Security and Privacy* (pp. 39-62).

www.irma-international.org/article/balanced-approach-for-hiding-sensitive-association-rules-in-data-sharing-environment/136365

Foundations for Cryptography

Dieter Gollmann (2000). *Internet and Intranet Security Management: Risks and Solutions* (pp. 163-184).

www.irma-international.org/chapter/foundations-cryptography/24601

Aspect-Oriented Analysis of Security in Distributed Virtual Environment

Li Yang, Raimund K. Egeand Lin Luo (2009). *Handbook of Research on Information Security and Assurance* (pp. 218-229).

www.irma-international.org/chapter/aspect-oriented-analysis-security-distributed/20652

Internet and Suicide

Dianne Currier (2007). *Encyclopedia of Information Ethics and Security* (pp. 384-390).

www.irma-international.org/chapter/internet-suicide/13500

Effect of Cultural Diversity in Communication on Project Success Within Multinational Construction Companies in the United Arab Emirates

Brian J. Galli (2019). *International Journal of Risk and Contingency Management* (pp. 39-65).

www.irma-international.org/article/effect-of-cultural-diversity-in-communication-on-project-success-within-multinational-construction-companies-in-the-united-arab-emirates/234433