Chapter XIX
A Web Metadata Based–Model for Information Quality Prediction

Ricardo Barros  
COPPE—Federal University of Rio de Janeiro, Brazil

Geraldo Xexéo  
COPPE—Federal University of Rio de Janeiro, Brazil

Wallace A. Pinheiro  
COPPE—Federal University of Rio de Janeiro, Brazil

Jano de Souza  
COPPE—Federal University of Rio de Janeiro, Brazil

ABSTRACT

Currently, in the Web environment, users have to deal with an enormous amount of information. In a Web search, they often receive useless, replicated, outdated, or false data, which, at first, they have no means to assess. Web search engines provide good examples of these problems: As reply from these mechanisms, users usually find links to replicated or conflicting information. Further, in these cases, information is spread out among heterogeneous and unrelated data sources, that normally present different information quality approaches. This chapter addresses those issues by proposing a Web Metadata-Based Model to evaluate and recommend Web pages based on their information quality, as predicted by their metadata. We adopt a fuzzy theory approach to obtain the values of quality dimensions from metadata values and to evaluate the quality of information, taking advantage of fuzzy logic’s ability to capture humans’ imprecise knowledge and deal with different concepts.

INTRODUCTION

Recently, the volume of information available to the users has increased enormously. At the Web, it is possible to search for information on an unlimited number of contexts and categories across a wide range of information environments, such as databases, application systems, electronic library systems, corporate intranets and the Internet as well. This information presents different
levels of quality, with original sources ranging from multi-national corporations to individuals with limited knowledge. With so much information available, quality has become an important discriminator when deciding which information to use and which to discard (Burgess, Gray, & Fiddian, 2004).

In this sense, the difficulties to identify, separate, and assess the quality of information have caused financial losses and compromised decision-making processes (Eckerson, 2002; English, 1999; Redman, 1998).

In spite of extensive discussion in literature, there is no consensus on an appropriate approach to improve the quality of information, as for the effectiveness of proposals and the expected benefits. However, there is a common sense that the effort to reach a good information quality standard must be high priority (Eckerson, 2002; Redman, 1998).

This chapter describes a model to explore the benefits of using Web metadata for information quality prediction. In our approach, we start from separated evaluations for each quality dimension and aggregate them regarding to several rules and context characteristics to obtain aggregated evaluations. We apply a fuzzy theory approach to obtain the values of quality dimensions from metadata values and to evaluate the quality of information, taking advantage of fuzzy logic’s ability to capture humans’ imprecise knowledge and deal with different concepts.

We also apply the fuzzy theory approach to identify user expectations about information quality (Xexéo, Belchior, & da Rocha, 1996; Yager, 1991; Zadeh, 1988).

We adopted an ontology represented by a UML² model to formalize, keep, and share the concepts and its instances used in all steps of the evaluation process³ to predict information quality.

There is a real example to illustrate how it has been used, taking the metadata update date, query time, update time, forward links, back-wards links, hubs, and authorities and applying them as basis to evaluate quality dimensions timeliness, completeness, and reputation in an economy context.

The remainder of this chapter is organized as follows. In the next section we present a background. Afterwards we demonstrate our proposal and show the architecture and some technical details as well. Finally, we delineate the expected results by an example, and we point out some future trends and conclude the chapter.

BACKGROUND

Brief Introduction to Fuzzy Logic

Fuzzy logic is also another extension realized in Boolean logic that may be considered a generalization of multi-valued logic. By modeling the uncertainties of natural language through concepts of partial truth – truth-values falling somewhere between completely true and completely false (Kantrowitz, Horstkotte, & Joslyn, 1997) – fuzzy logic deals with such values through fuzzy sets in the interval [0,1]. These characteristics allow fuzzy logic to manipulate real-world objects that possess imprecise limits. Utilizing fuzzy predicates (old, new, high, etc.), fuzzy quantifiers (many, few, almost all, etc.), fuzzy truth-values (completely true, more or less true) (Dubois & Prade, 1991) and generalizing the meaning of connectors and logical operators, fuzzy logic is seen as a means of approximate reasoning (Grauel, 1999).

It was introduced by Dr. Lotfi Zadeh of UC/ Berkeley in the 1960’s as a way to model the uncertainty of natural language. Zadeh (1965) says that rather than regarding fuzzy theory as a single theory, we should regard the process of “fuzzification” as a methodology to generalize any specific theory from a crisp (discrete) to a continuous (fuzzy) form. Thus, recently researchers have also introduced “fuzzy calculus”, “fuzzy differential equations”, “fuzzy systems”, “fuzzy
18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/chapter/web-metadata-based-model-information/21981?camid=4v1


www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Integrating Accessibility Evaluation into Web Engineering Processes
www.igi-global.com/chapter/integrating-accessibility-evaluation-into-web/53067?camid=4v1a

Design and Implementation of Wireless Voltage Monitoring System Based on Zigbee
www.igi-global.com/article/design-and-implementation-of-wireless-voltage-monitoring-system-based-on-zigbee/182267?camid=4v1a

An Unsupervised Approach for Determining Link Specifications
www.igi-global.com/article/an-unsupervised-approach-for-determining-link-specifications/209723?camid=4v1a

Evaluating Social Interaction and Support Methods over Time
www.igi-global.com/chapter/evaluating-social-interaction-and-support-methods-over-time/137398?camid=4v1a