

## Chapter III

# Web Usage Mining for Ontology Management

**Brigitte Trousse**

*INRIA Sophia Antipolois, France*

**Marie-Aude Aufaure**

*INRIA Sophia and Supélec, France*

**Bénédicte Le Grand**

*Laboratoire d'Informatique de Paris 6, France*

**Yves Lechevallier**

*INRIA Rocquencourt, France*

**Florent Massegia**

*INRIA Sophia Antipolois, France*

### **ABSTRACT**

*This chapter proposes an original approach for ontology management in the context of Web-based information systems. Our approach relies on the usage analysis of the chosen Web site, in addition to the existing approaches based on Web pages content analysis. Our methodology is based on knowledge discovery techniques mainly from HTTP Web logs and aims to confronting the discovered knowledge in terms of usage with the existing ontology in order to propose new relations between concepts. We illustrate our approach on a Web site provided by local French tourism authorities (related to Metz city) with the use of clustering and sequential patterns discovery methods. One major contribution of this chapter is, thus, the application of usage analysis to support ontology evolution and/or Web site reorganization.*

## INTRODUCTION

Finding relevant information on the Web has become a real challenge. This is partly due to the volume of available data and the lack of structure in many Web sites. However, information retrieval may also be difficult in well-structured sites, such as those of tourist offices. This is not only due to the volume of data, but also the way information is organized, as it does not necessarily meet Internet users' expectations. Mechanisms are necessary to enhance their understanding of visited sites.

Local tourism authorities have developed Web sites in order to promote tourism and to offer services to citizens. However, this information is scattered and unstructured and thus does not match tourists' expectations. Our work aims at to provide a solution to this problem by:

- Using an existing or semiautomatically built ontology intended to enhance information retrieval.
- Identifying Web users' profiles through an analysis of their visits employing Web usage mining methods, in particular automatic classification and sequential pattern mining techniques.
- Updating the ontology with extracted knowledge. We will study the impact of the visit profiles on the ontology. In particular, we will propose to update Web sites by adapting their structure to a given visit profile. For example, we will propose to add new hyperlinks in order to be consistent with the new ontology.

The first task relies on Web structure and content mining, while the second and the third are deduced from usage analysis. A good structure of information will allow us to extract knowledge from log files (traces of visit on the Web), extracted knowledge will help us update Web sites' ontology according to tourists' expectations. Local tourism authorities will thus be able to use these models

and check whether their tourism policy matches tourists' behavior. This is essential in the domain of tourism, which is highly competitive. Moreover, the Internet is widely available and tourists may even navigate pages on the Web through wireless connections. It is therefore necessary to develop techniques and tools in order to help them find relevant information easily and quickly.

In the future, we will propose to personalize the display, in order to adapt it to each individual's preferences according to his profile. This will be achieved through the use of weights on ontology concepts and relations extracted from usage analysis.

This chapter is composed of five sections. The first section presents the main definitions useful for the comprehension. In the second, we briefly describe the state-of-the-art in ontology management and Web mining. Then, we propose our methodology mainly based on Web usage mining for supporting ontology management and Web site reorganization. Finally, before concluding, we illustrate the proposed methodology in the tourism domain.

## BACKGROUND

This section presents the context of our work and is divided into two subsections. We first provide the definitions of the main terms we use in this chapter, then we study the state-of-the-art techniques related to ontology management and to Web mining.

### Definitions

- *Web mining* is a knowledge discovery process from databases (called KDD) applied to Web data. Web mining can extract patterns from various types of data; three areas of Web mining are often distinguished: Web content mining, Web structure mining and

26 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-usage-mining-ontology-management/7571](http://www.igi-global.com/chapter/web-usage-mining-ontology-management/7571)

## Related Content

---

### Measuring Semantic-Based Structural Similarity in Multi-Relational Networks

Yunchuan Sun, Rongfang Bie and Junsheng Zhang (2016). *International Journal of Data Warehousing and Mining* (pp. 20-33).

[www.irma-international.org/article/measuring-semantic-based-structural-similarity-in-multi-relational-networks/143713](http://www.irma-international.org/article/measuring-semantic-based-structural-similarity-in-multi-relational-networks/143713)

### Effectiveness of Fuzzy Classifier Rules in Capturing Correlations between Genes

Mohammad Khabbaz, Keivan Kianmehr, Mohammad Alshalalfa and Reda Alhajj (2008). *International Journal of Data Warehousing and Mining* (pp. 62-83).

[www.irma-international.org/article/effectiveness-fuzzy-classifier-rules-capturing/1818](http://www.irma-international.org/article/effectiveness-fuzzy-classifier-rules-capturing/1818)

### NewSum: "N-Gram Graph"-Based Summarization in the Real World

George Giannakopoulos, George Kiomourtzis and Vangelis Karkaletsis (2014). *Innovative Document Summarization Techniques: Revolutionizing Knowledge Understanding* (pp. 205-230).

[www.irma-international.org/chapter/newsum/96746](http://www.irma-international.org/chapter/newsum/96746)

### Feature Selection for the Promoter Recognition and Prediction Problem

George Potamias and Alexandros Kanterakis (2007). *International Journal of Data Warehousing and Mining* (pp. 60-78).

[www.irma-international.org/article/feature-selection-promoter-recognition-prediction/1790](http://www.irma-international.org/article/feature-selection-promoter-recognition-prediction/1790)

### Tasks, Approaches, and Avenues of Opinion Mining, Sentiment Analysis, and Emotion Analysis: Opinion Mining and Extents

Amira M. Idrees, Fatma Gamal Eldin, Amr Mansour Mohsen and Hesham Ahmed Hassan (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 77-115).

[www.irma-international.org/chapter/tasks-approaches-and-avenues-of-opinion-mining-sentiment-analysis-and-emotion-analysis/308481](http://www.irma-international.org/chapter/tasks-approaches-and-avenues-of-opinion-mining-sentiment-analysis-and-emotion-analysis/308481)