

# Short Book Reviews

## **Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage**

Zdravko Markov, Daniel T. Larose

Wiley, 2007, xvi + 218 pages, £ 38.95 / € 56.00, hardcover

ISBN: 978-0-471-66655-4

### *Table of contents*

#### Part I: Web structure mining

1. Information retrieval and web search
2. Hyperlink-based ranking

#### Part II: Web content mining

3. Clustering
4. Evaluating clustering
5. Classification

#### Part III: Web usage mining

6. Introduction to web usage mining
7. Preprocessing for web usage mining
8. Exploratory data analysis for web usage mining
9. Modeling for web usage mining:  
clustering, association, and classification

*Readership:* Graduate and advanced undergraduate students in computer science and engineering, web designers and researchers.

This book is the third volume in a data mining series, with the two previous ones, authored by Daniel Larose, entitled *Discovering Knowledge in Data: An introduction to Data Mining* and *Data Mining Methods and Models*. The book stands independently of the earlier two, addressing the particular data mining topics of the web. It is divided into three parts concerned with, respectively, mining the structure, content, and usage of the web.

Web mining is an increasingly important sub-domain of data mining. It has multiple aspects – and, of course, intersects heavily with other application domains of data mining. This book aims to explain the operation of the relevant data mining algorithms and tools by leading the reader through applications of them to small data sets, and then illustrating them by applying the applications to real large data sets. The authors have used the Weka software system (a public domain data analysis collection) and SPSS Clementine to illustrate. An extensive range of exercises is included.

The book is supported by a website. Although not complete at the time of the review, the aim is to include the data sets used in the book and links to other data mining resources. For faculty who adopt the book for teaching, there is intended to be a further restricted-access website which includes solutions to the exercises, Powerpoint presentations of each chapter, sample data mining projects, and other resources.

Overall, the book is highly accessible and clearly presented. In my view it provides an excellent introduction to mining the web.

David J. Hand  
Mathematics Department, Imperial College  
London SW7 2AZ, UK  
E-mail: d.j.hand@imperial.ac.uk