Briefly Noted

Text Mining: Applications and Theory

Michael W. Berry* and Jacob Kogan† (editors)
(*University of Tennessee and †University of Maryland, Baltimore County)

Hoboken, NJ: Wiley, 2010, xiv+207 pp; hardbound, ISBN 978-0-470-74982-1, £55.00, €66.00

This book is the fully revised proceedings of a 2009 one-day workshop on text mining. Collectively, the ten contributions span several major topic areas in text mining, divided into three parts: keyword extraction, classification, and clustering; anomaly and trend detection; and text streams. Generally, the articles in the first part are technical reports, and most of those in the second and third parts are research reviews.

Compared with earlier books on the topic (Berry 2003; Weiss et al. 2005; Srivastava and Sahami 2009), Chapter 6, entitled "Survey of text visualization techniques" is really new, enlarging text mining to a new field of scenario discovery. The authors, Andrey A. Puretskiy, Gregory L. Shutt, and Michael W. Berry, explore several visual techniques and describe specific examples of software that uses them. It's a demand-oriented study from the perspective of users' needs or purposes. Every technique is concentrated on one certain user's purpose. For example, the tag clouds technique is introduced and shown to be highly useful in the area of "quick, complete, and graphical summary of a large document"' (page 108). One of the visual post-processing tools tailored for specific text mining packages, named FutureLens, is discussed in great detail in this chapter.

Not all of the research results in this book are good, however. For example, in Chapter 1, "Automatic keyword extraction from individual documents," by Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley, a relatively high rate of keyword extraction precision should be important in information retrieval (IR) systems, but the results shown include a precision of only 67%. All in all, this book is a good presentation of the state-of-the-art algorithms for text mining from both the academic and industrial perspectives, and a useful volume for practitioners and students in computer science, natural language processing, bioinformatics, and engineering who wish to use text-mining techniques.—*Zhang Xiaojun, Shaanxi Normal University*

References

Berry, Michael W., editor. 2003. *Survey of Text Mining: Clustering, Classification, and Retrieval*. Springer, New York.

Srivastava, Ashok N. and Mehran Sahami, editors. 2009. *Text Mining: Classification, Clustering, and Applications*. CRC Press, Boca Raton, FL.

Weiss, Sholom M., Nitin Indurkhya, Tong Zhang, and Fred J. Damerau. 2005. *Text Mining: Predictive Methods for Analyzing Unstructured Information*. Springer, New York.