One consequence of the pervasive use of computers is that most documents originate in digital form. Text mining—the process of searching, retrieving, and analyzing unstructured, natural-language text—is concerned with how to exploit the textual data embedded in these documents.

Text Mining presents a comprehensive introduction and overview of the field, integrating related topics (such as artificial intelligence and knowledge discovery and data mining) and providing practical advice on how readers can use text-mining methods to analyze their own data. Emphasizing predictive methods, the book unifies all key areas in text mining: preprocessing, text categorization, information search and retrieval, clustering of documents, and information extraction. In addition, it identifies emerging directions for those looking to do research in the area. Some background in data mining is beneficial, but not essential.

Topics and features:

- Presents a comprehensive and easy-to-read introduction to text mining
- Explores the application and utility of the methods, as well as the optimal techniques for specific scenarios
- Provides several descriptive case studies that take readers from problem description to system deployment in the real world
- Uses methods that rely on basic statistical techniques, thus allowing for relevance to all languages (not just English)
- Includes access to downloadable software (runs on any computer), as well as useful chapter-ending historical and bibliographical remarks, a detailed bibliography, and subject and author indexes

This authoritative and highly accessible text, written by a team of authorities on text mining, develops the foundation concepts, principles, and methods needed to expand beyond structured, numeric data to automated mining of text samples. Researchers, professionals, and advanced undergraduates and graduates with work and interests in data mining, machine learning, databases, and computational linguistics will find the work an essential resource.

ISBN 0-387-95433-3
90000
9 780387 954332

springeronline.com