



Mining the Web: Discovering Knowledge from Hypertext Data

By Soumen Chakrabarti

Download now

Read Online ➔

Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti

Mining the Web: Discovering Knowledge from Hypertext Data is the first book devoted entirely to techniques for producing knowledge from the vast body of unstructured Web data. Building on an initial survey of infrastructural issues?including Web crawling and indexing?Chakrabarti examines low-level machine learning techniques as they relate specifically to the challenges of Web mining. He then devotes the final part of the book to applications that unite infrastructure and analysis to bring machine learning to bear on systematically acquired and stored data. Here the focus is on results: the strengths and weaknesses of these applications, along with their potential as foundations for further progress. From Chakrabarti's work?painstaking, critical, and forward-looking?readers will gain the theoretical and practical understanding they need to contribute to the Web mining effort.

- * A comprehensive, critical exploration of statistics-based attempts to make sense of Web Mining.
- * Details the special challenges associated with analyzing unstructured and semi-structured data.
- * Looks at how classical Information Retrieval techniques have been modified for use with Web data.
- * Focuses on today's dominant learning methods: clustering and classification, hyperlink analysis, and supervised and semi-supervised learning.
- * Analyzes current applications for resource discovery and social network analysis.
- * An excellent way to introduce students to especially vital applications of data mining and machine learning technology.

↓ [Download Mining the Web: Discovering Knowledge from Hyperte...pdf](#)

📖 [Read Online Mining the Web: Discovering Knowledge from Hyper...pdf](#)

Mining the Web: Discovering Knowledge from Hypertext Data

By Soumen Chakrabarti

Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti

Mining the Web: Discovering Knowledge from Hypertext Data is the first book devoted entirely to techniques for producing knowledge from the vast body of unstructured Web data. Building on an initial survey of infrastructural issues?including Web crawling and indexing?Chakrabarti examines low-level machine learning techniques as they relate specifically to the challenges of Web mining. He then devotes the final part of the book to applications that unite infrastructure and analysis to bring machine learning to bear on systematically acquired and stored data. Here the focus is on results: the strengths and weaknesses of these applications, along with their potential as foundations for further progress. From Chakrabarti's work?painstaking, critical, and forward-looking?readers will gain the theoretical and practical understanding they need to contribute to the Web mining effort.

- * A comprehensive, critical exploration of statistics-based attempts to make sense of Web Mining.
- * Details the special challenges associated with analyzing unstructured and semi-structured data.
- * Looks at how classical Information Retrieval techniques have been modified for use with Web data.
- * Focuses on today's dominant learning methods: clustering and classification, hyperlink analysis, and supervised and semi-supervised learning.
- * Analyzes current applications for resource discovery and social network analysis.
- * An excellent way to introduce students to especially vital applications of data mining and machine learning technology.

Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti Bibliography

- Sales Rank: #612512 in Books
- Brand: Brand: Morgan Kaufmann
- Published on: 2002-10-23
- Original language: English
- Number of items: 1
- Dimensions: 9.50" h x 7.50" w x 1.00" l, 1.66 pounds
- Binding: Hardcover
- 344 pages

 [Download Mining the Web: Discovering Knowledge from Hyperte ...pdf](#)

 [Read Online Mining the Web: Discovering Knowledge from Hyper ...pdf](#)

Editorial Review

Review

"...solid and beneficial to readers interested in Web data mining, especially those interested in the details of algorithmic implementation." = Bernard J. Jansen, *Information Processing & Management*

"The treatment is systematic, comprehensive and in-depth, yet very lucid and accessible to a wide range of Web technology developers. The author's insights and depth of knowledge as one of the pioneering researchers on hypertext information mining and retrieval are also evident in the extensive and useful bibliographic notes provided at the end of each chapter..." - Professor Joydeep Ghosh, University of Texas, Austin

"The author has done the community a great service by synthesizing all the important work in this field into an excellent book, which introduces fairly sophisticated material in an easy-to-read manner. This book for the first time, makes it possible to offer Web Mining as a real course." - Professor Jaideep Srivastava, University of Minnesota

" *Mining the Web: Discovering Knowledge from Hypertext from Hypertext Data*, by Soumen Chakrabarti, focuses extensively on building a better search engine crawler...Chakrabarti's book begins with a discussion of search engine crawlers in a chapter titled "Crawling the Web." The discussion in this chapter is technical and detailed. Readers learn about features such as the robots.txt file that can be written in a certain way to stop crawlers from visiting a page...The most interesting part of the book is perhaps Chapter 7, "Social Network Analysis." In this chapter, the author presents the most famous search engine algorithms (e.g., PageRank, HITS, SALSA)." - Journal of Marketing Research, Sandeep Krishnamurthy

"All in all this is an excellent book. I enjoyed the book and highly recommend it as a textbook for web data mining classes at graduate or senior undergraduate levels. Chakrabarti has a rich vocabulary and is a gifted writer. I bet he will write new, good books in the future, and he should. I look forward to them." - Fazli Can - Miami University

From the Back Cover

Mining the Web: Discovering Knowledge from Hypertext Data is the first book devoted entirely to techniques for extracting and producing knowledge from the vast body of unstructured Web data. Building on an initial survey of infrastructural issues?including Web crawling and indexing?Chakrabarti examines machine learning techniques as they relate specifically to the challenges of Web mining and provides applications of machine learning to systematically acquire, store, and analyze data. Here the focus is on results: the strengths and weaknesses of these applications, along with their potential as foundations for further progress toward a Web that is more aware of content semantics. This thorough and forward-looking book gives the theoretical and practical foundations you need to build innovative applications for mining the Web.

Features

- A comprehensive, critical exploration of statistics-based attempts to make sense of Web data.
- Details the special challenges associated with analyzing unstructured and semi-structured data.
- Looks at how classical Information Retrieval techniques have been modified for use with Web data.

- Focuses on today's dominant learning methods: clustering and classification, hyperlink analysis, and supervised and semi-supervised learning.
- Analyzes current applications for resource discovery and social network analysis.
- An excellent way to introduce students to especially vital applications of data mining and machine learning technology.

About the Author

Soumen Chakrabarti is assistant Professor in Computer Science and Engineering at the Indian Institute of Technology, Bombay. Prior to joining IIT, he worked on hypertext databases and data mining at IBM Almaden Research Center. He has developed three systems and holds five patents in this area. Chakrabarti has served as a vice-chair and program committee member for many conferences, including WWW, SIGIR, ICDE, and KDD, and as a guest editor of the IEEE TKDE special issue on mining and searching the Web. His work on focused crawling received the Best Paper award at the 8th International World Wide Web Conference (1999). He holds a Ph.D. from the University of California, Berkeley.

Users Review

From reader reviews:

Joel Jones:

Have you spare time for just a day? What do you do when you have much more or little spare time? Yes, you can choose the suitable activity regarding spend your time. Any person spent their very own spare time to take a wander, shopping, or went to often the Mall. How about open or even read a book eligible Mining the Web: Discovering Knowledge from Hypertext Data? Maybe it is for being best activity for you. You recognize beside you can spend your time along with your favorite's book, you can smarter than before. Do you agree with their opinion or you have some other opinion?

Scott Settle:

This Mining the Web: Discovering Knowledge from Hypertext Data book is not ordinary book, you have after that it the world is in your hands. The benefit you obtain by reading this book is usually information inside this reserve incredible fresh, you will get info which is getting deeper you read a lot of information you will get. This specific Mining the Web: Discovering Knowledge from Hypertext Data without we understand teach the one who looking at it become critical in thinking and analyzing. Don't become worry Mining the Web: Discovering Knowledge from Hypertext Data can bring whenever you are and not make your handbag space or bookshelves' become full because you can have it in the lovely laptop even mobile phone. This Mining the Web: Discovering Knowledge from Hypertext Data having excellent arrangement in word and layout, so you will not really feel uninterested in reading.

Joseph Gabriel:

Here thing why this particular Mining the Web: Discovering Knowledge from Hypertext Data are different and reputable to be yours. First of all studying a book is good but it really depends in the content of the usb ports which is the content is as scrumptious as food or not. Mining the Web: Discovering Knowledge from Hypertext Data giving you information deeper and different ways, you can find any publication out there but

there is no publication that similar with Mining the Web: Discovering Knowledge from Hypertext Data. It gives you thrill reading journey, its open up your current eyes about the thing that happened in the world which is might be can be happened around you. You can actually bring everywhere like in park your car, café, or even in your way home by train. For anyone who is having difficulties in bringing the branded book maybe the form of Mining the Web: Discovering Knowledge from Hypertext Data in e-book can be your alternate.

Heather Garcia:

The actual book Mining the Web: Discovering Knowledge from Hypertext Data has a lot of information on it. So when you check out this book you can get a lot of advantage. The book was compiled by the very famous author. This articles author makes some research just before write this book. This particular book very easy to read you will get the point easily after perusing this book.

**Download and Read Online Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti
#ZX1BORSKEJY**

Read Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti for online ebook

Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti Free PDF d0wnl0ad, audio books, books to read, good books to read, cheap books, good books, online books, books online, book reviews epub, read books online, books to read online, online library, greatbooks to read, PDF best books to read, top books to read Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti books to read online.

Online Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti ebook PDF download

Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti Doc

Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti Mobipocket

Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti EPub