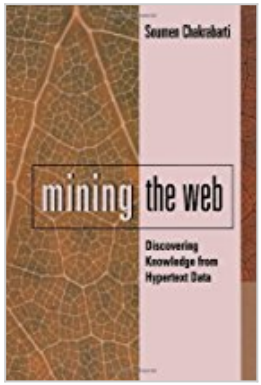

Mining the Web: Discovering Knowledge from Hypertext Data

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

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

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 on 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 sytematically 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:

Erica Rawlins:

What do you consider book? It is just for students since they're still students or it for all people in the world, the particular best subject for that? Simply you can be answered for that query above. Every person has distinct personality and hobby for each other. Don't to be compelled someone or something that they don't desire do that. You must know how great along with important the book Mining the Web: Discovering Knowledge from Hypertext Data. All type of book could you see on many solutions. You can look for the internet methods or other social media.

Donald Jackson:

The book Mining the Web: Discovering Knowledge from Hypertext Data will bring you to definitely the new experience of reading the book. The author style to clarify the idea is very unique. In the event you try to find new book to see, this book very suitable to you. The book Mining the Web: Discovering Knowledge from Hypertext Data is much recommended to you to see. You can also get the e-book from your official web site, so you can more readily to read the book.

Harriette Corwin:

Reading can called imagination hangout, why? Because if you are reading a book mainly book entitled Mining the Web: Discovering Knowledge from Hypertext Data your mind will drift away trough every dimension, wandering in most aspect that maybe unidentified for but surely can be your mind friends. Imaging each word written in a publication then become one web form conclusion and explanation this maybe you never get previous to. The Mining the Web: Discovering Knowledge from Hypertext Data giving you one more experience more than blown away your thoughts but also giving you useful info for your better life on this era. So now let us teach you the relaxing pattern is your body and mind will be pleased when you are finished reading it, like winning a. Do you want to try this extraordinary investing spare time activity?

Monika Cunniff:

Beside this specific Mining the Web: Discovering Knowledge from Hypertext Data in your phone, it may give you a way to get closer to the new knowledge or info. The information and the knowledge you may got here is fresh in the oven so don't become worry if you feel like an previous people live in narrow commune. It is good thing to have Mining the Web: Discovering Knowledge from Hypertext Data because this book offers to your account readable information. Do you often have book but you don't get what it's facts concerning. Oh come on, that will not happen if you have this in your hand. The Enjoyable agreement here cannot be questionable, such as treasuring beautiful island. So do you still want to miss the item? Find this book and read it from at this point!

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