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

 Download Mining the Web: Discovering Knowledge from Hyperte ...pdf

 Read Online Mining the Web: Discovering Knowledge from Hyper ...pdf

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

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:

John Charlie:

The reason? Because this Mining the Web: Discovering Knowledge from Hypertext Data is an unordinary book that the inside of the publication waiting for you to snap the item but latter it will surprise you with the secret that inside. Reading this book close to it was fantastic author who write the book in such incredible way makes the content inside of easier to understand, entertaining way but still convey the meaning fully. So , it is good for you for not hesitating having this any more or you going to regret it. This amazing book will give you a lot of benefits than the other book include such as help improving your skill and your critical thinking means. So , still want to hold up having that book? If I have been you I will go to the reserve store hurriedly.

Charles Whittaker:

The book untitled Mining the Web: Discovering Knowledge from Hypertext Data contain a lot of information on the idea. The writer explains your ex idea with easy means. The language is very easy to understand all the people, so do not worry, you can easy to read the idea. The book was written by famous author. The author brings you in the new era of literary works. It is easy to read this book because you can please read on your smart phone, or product, so you can read the book within anywhere and anytime. If you want to buy the e-book, you can start their official web-site as well as order it. Have a nice go through.

Tia Rosario:

Beside this Mining the Web: Discovering Knowledge from Hypertext Data in your phone, it can give you a way to get closer to the new knowledge or data. The information and the knowledge you can got here is fresh from your oven so don't possibly be worry if you feel like an outdated people live in narrow town. 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 oftentimes have book but you do not get what it's facts

concerning. Oh come on, that will not end up to happen if you have this in the hand. The Enjoyable blend here cannot be questionable, just like treasuring beautiful island. So do you still want to miss the item? Find this book and read it from currently!

Norma Barnes:

That reserve can make you to feel relax. This specific book Mining the Web: Discovering Knowledge from Hypertext Data was colorful and of course has pictures on the website. As we know that book Mining the Web: Discovering Knowledge from Hypertext Data has many kinds or type. Start from kids until teenagers. For example Naruto or Detective Conan you can read and believe you are the character on there. Therefore , not at all of book usually are make you bored, any it can make you feel happy, fun and loosen up. Try to choose the best book to suit your needs and try to like reading that.

Download and Read Online Mining the Web: Discovering Knowledge from Hypertext Data By Soumen Chakrabarti #PHI12Z7OMWS

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