



An investigation of web crawler behavior: characterization and metrics

Marios D. Dikaiakos^{a,*}, Athena Stassopoulou^b, Loizos Papageorgiou^a

^aDepartment of Computer Science, University of Cyprus, P.O. Box 20537, Nicosia 1678, Cyprus

^bDepartment of Computer Science, Asercollege, P.O. Box 24005, Nicosia, Cyprus

Received 22 July 2004; revised 9 January 2005; accepted 11 January 2005

Available online 25 February 2005

Abstract

In this paper, we present a characterization study of search-engine crawlers. For the purposes of our work, we use Web-server access logs from five academic sites in three different countries. Based on these logs, we analyze the activity of different crawlers that belong to five search engines: Google, AltaVista, Inktomi, FastSearch and CiteSeer. We compare crawler behavior to the characteristics of the general World-Wide Web traffic and to general characterization studies. We analyze crawler requests to derive insights into the behavior and strategy of crawlers. We propose a set of simple metrics that describe qualitative characteristics of crawler behavior, viz. a crawler's preference on resources of a particular format, its frequency of visits on a Web site, and the pervasiveness of its visits to a particular site. To the best of our knowledge, this is the first extensive and in depth characterization of search-engine crawlers. Our results and observations provide useful insights into crawler behavior and serve as basis of our ongoing work on the automatic detection of Web crawlers.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Web characterization; Crawlers

1. Introduction

A *Web crawler* is a program that traverses the hypertext structure of the Web, starting from a ‘seed’ list of hyperdocuments and recursively retrieving documents accessible from that list [7,13,29]. Web crawlers are also referred to as robots, wanderers, or spiders. All major search engines employ powerful crawlers that traverse the Internet continuously, trying to discover and retrieve as many Web pages as possible. More recently, crawler systems have been used also as tools for focused crawling, for shopbot implementation, and for supporting added-value services on the Web (portals, personalized and mobile services, etc.) [14,19,29]. As a consequence, the number and the variety of active robots operating on the Internet increases continuously, resulting to a noticeable impact on WWW traffic and Web-server activity.

In this paper, we seek to characterize the behavior of crawlers that belong to popular Web-search engines. In particular, we investigate statistically various properties of crawler-induced HTTP traffic, such as the distribution of HTTP requests and reply-codes, the type and size of resources sought and retrieved, the distribution of crawler requests across a Web site, the frequency and pattern of crawler re-visits, and the temporal characteristics of crawler activity. We use these analyses to discover, characterize and compare the capabilities of different crawlers regarding their usage of caching, their capability to detect and avoid broken or dead links, their success in discovering Web-site resources, etc. Furthermore, we propose simple metrics that can be used to describe concisely aspects of crawling strategy and to compare basic crawler characteristics. Characterizing crawler activity is important as it enables researchers to: (i) estimate the impact of robots on the workload and performance of Web servers; (ii) investigate the contribution of crawlers to WWW traffic; (iii) discover and compare the strategies employed by different crawlers to reap resources from the Web, and (iv) model the activity of robots to produce synthetic crawler workloads for simulation studies. Finally, characterization can be the basis for the automatic detection of robots.

For the purposes of our study, we focus on the characterization of five crawlers that belong to well-known

* Corresponding author. Tel.: +357 22862700; fax: +357 22892701.
E-mail addresses: md@ucy.ac.cy (M.D. Dikaiakos), stassopoulou.a@asercollege.ac.cy (A. Stassopoulou), loipap@ucy.ac.cy (L. Papageorgiou).

An Investigation Of Web Crawler Behavior Characterization

JL Elias



An Investigation Of Web Crawler Behavior Characterization:

Discover tales of courage and bravery in Crafted by is empowering ebook, Stories of Fearlessness: **An Investigation Of Web Crawler Behavior Characterization** . In a downloadable PDF format (PDF Size: *), this collection inspires and motivates. Download now to witness the indomitable spirit of those who dared to be brave.

https://socketapi.adit.com/book/uploaded-files/fetch.php/cd_rates_sat_practice_review.pdf

Table of Contents An Investigation Of Web Crawler Behavior Characterization

1. Understanding the eBook An Investigation Of Web Crawler Behavior Characterization
 - The Rise of Digital Reading An Investigation Of Web Crawler Behavior Characterization
 - Advantages of eBooks Over Traditional Books
2. Identifying An Investigation Of Web Crawler Behavior Characterization
 - Exploring Different Genres
 - Considering Fiction vs. Non-Fiction
 - Determining Your Reading Goals
3. Choosing the Right eBook Platform
 - Popular eBook Platforms
 - Features to Look for in an An Investigation Of Web Crawler Behavior Characterization
 - User-Friendly Interface
4. Exploring eBook Recommendations from An Investigation Of Web Crawler Behavior Characterization
 - Personalized Recommendations
 - An Investigation Of Web Crawler Behavior Characterization User Reviews and Ratings
 - An Investigation Of Web Crawler Behavior Characterization and Bestseller Lists
5. Accessing An Investigation Of Web Crawler Behavior Characterization Free and Paid eBooks
 - An Investigation Of Web Crawler Behavior Characterization Public Domain eBooks
 - An Investigation Of Web Crawler Behavior Characterization eBook Subscription Services
 - An Investigation Of Web Crawler Behavior Characterization Budget-Friendly Options
6. Navigating An Investigation Of Web Crawler Behavior Characterization eBook Formats

- ePub, PDF, MOBI, and More
 - An Investigation Of Web Crawler Behavior Characterization Compatibility with Devices
 - An Investigation Of Web Crawler Behavior Characterization Enhanced eBook Features
7. Enhancing Your Reading Experience
 - Adjustable Fonts and Text Sizes of An Investigation Of Web Crawler Behavior Characterization
 - Highlighting and Note-Taking An Investigation Of Web Crawler Behavior Characterization
 - Interactive Elements An Investigation Of Web Crawler Behavior Characterization
 8. Staying Engaged with An Investigation Of Web Crawler Behavior Characterization
 - Joining Online Reading Communities
 - Participating in Virtual Book Clubs
 - Following Authors and Publishers An Investigation Of Web Crawler Behavior Characterization
 9. Balancing eBooks and Physical Books An Investigation Of Web Crawler Behavior Characterization
 - Benefits of a Digital Library
 - Creating a Diverse Reading Collection An Investigation Of Web Crawler Behavior Characterization
 10. Overcoming Reading Challenges
 - Dealing with Digital Eye Strain
 - Minimizing Distractions
 - Managing Screen Time
 11. Cultivating a Reading Routine An Investigation Of Web Crawler Behavior Characterization
 - Setting Reading Goals An Investigation Of Web Crawler Behavior Characterization
 - Carving Out Dedicated Reading Time
 12. Sourcing Reliable Information of An Investigation Of Web Crawler Behavior Characterization
 - Fact-Checking eBook Content of An Investigation Of Web Crawler Behavior Characterization
 - Distinguishing Credible Sources
 13. Promoting Lifelong Learning
 - Utilizing eBooks for Skill Development
 - Exploring Educational eBooks
 14. Embracing eBook Trends
 - Integration of Multimedia Elements
 - Interactive and Gamified eBooks

An Investigation Of Web Crawler Behavior Characterization Introduction

Free PDF Books and Manuals for Download: Unlocking Knowledge at Your Fingertips In today's fast-paced digital age, obtaining valuable knowledge has become easier than ever. Thanks to the internet, a vast array of books and manuals are now available for free download in PDF format. Whether you are a student, professional, or simply an avid reader, this treasure trove of downloadable resources offers a wealth of information, conveniently accessible anytime, anywhere. The advent of online libraries and platforms dedicated to sharing knowledge has revolutionized the way we consume information. No longer confined to physical libraries or bookstores, readers can now access an extensive collection of digital books and manuals with just a few clicks. These resources, available in PDF, Microsoft Word, and PowerPoint formats, cater to a wide range of interests, including literature, technology, science, history, and much more. One notable platform where you can explore and download free An Investigation Of Web Crawler Behavior Characterization PDF books and manuals is the internet's largest free library. Hosted online, this catalog compiles a vast assortment of documents, making it a veritable goldmine of knowledge. With its easy-to-use website interface and customizable PDF generator, this platform offers a user-friendly experience, allowing individuals to effortlessly navigate and access the information they seek. The availability of free PDF books and manuals on this platform demonstrates its commitment to democratizing education and empowering individuals with the tools needed to succeed in their chosen fields. It allows anyone, regardless of their background or financial limitations, to expand their horizons and gain insights from experts in various disciplines. One of the most significant advantages of downloading PDF books and manuals lies in their portability. Unlike physical copies, digital books can be stored and carried on a single device, such as a tablet or smartphone, saving valuable space and weight. This convenience makes it possible for readers to have their entire library at their fingertips, whether they are commuting, traveling, or simply enjoying a lazy afternoon at home. Additionally, digital files are easily searchable, enabling readers to locate specific information within seconds. With a few keystrokes, users can search for keywords, topics, or phrases, making research and finding relevant information a breeze. This efficiency saves time and effort, streamlining the learning process and allowing individuals to focus on extracting the information they need. Furthermore, the availability of free PDF books and manuals fosters a culture of continuous learning. By removing financial barriers, more people can access educational resources and pursue lifelong learning, contributing to personal growth and professional development. This democratization of knowledge promotes intellectual curiosity and empowers individuals to become lifelong learners, promoting progress and innovation in various fields. It is worth noting that while accessing free An Investigation Of Web Crawler Behavior Characterization PDF books and manuals is convenient and cost-effective, it is vital to respect copyright laws and intellectual property rights. Platforms offering free downloads often operate within legal boundaries, ensuring that the materials they provide are either in the public domain or authorized for distribution. By adhering to copyright laws, users can enjoy the

benefits of free access to knowledge while supporting the authors and publishers who make these resources available. In conclusion, the availability of An Investigation Of Web Crawler Behavior Characterization free PDF books and manuals for download has revolutionized the way we access and consume knowledge. With just a few clicks, individuals can explore a vast collection of resources across different disciplines, all free of charge. This accessibility empowers individuals to become lifelong learners, contributing to personal growth, professional development, and the advancement of society as a whole. So why not unlock a world of knowledge today? Start exploring the vast sea of free PDF books and manuals waiting to be discovered right at your fingertips.

FAQs About An Investigation Of Web Crawler Behavior Characterization Books

How do I know which eBook platform is the best for me? Finding the best eBook platform depends on your reading preferences and device compatibility. Research different platforms, read user reviews, and explore their features before making a choice. Are free eBooks of good quality? Yes, many reputable platforms offer high-quality free eBooks, including classics and public domain works. However, make sure to verify the source to ensure the eBook credibility. Can I read eBooks without an eReader? Absolutely! Most eBook platforms offer web-based readers or mobile apps that allow you to read eBooks on your computer, tablet, or smartphone. How do I avoid digital eye strain while reading eBooks? To prevent digital eye strain, take regular breaks, adjust the font size and background color, and ensure proper lighting while reading eBooks. What the advantage of interactive eBooks? Interactive eBooks incorporate multimedia elements, quizzes, and activities, enhancing the reader engagement and providing a more immersive learning experience. An Investigation Of Web Crawler Behavior Characterization is one of the best book in our library for free trial. We provide copy of An Investigation Of Web Crawler Behavior Characterization in digital format, so the resources that you find are reliable. There are also many Ebooks of related with An Investigation Of Web Crawler Behavior Characterization. Where to download An Investigation Of Web Crawler Behavior Characterization online for free? Are you looking for An Investigation Of Web Crawler Behavior Characterization PDF? This is definitely going to save you time and cash in something you should think about.

Find An Investigation Of Web Crawler Behavior Characterization :

[cd rates sat practice review](#)

[side hustle ideas top warranty](#)

[**weekly ad buy online returns**](#)

[irs refund status today tutorial](#)

[walking workout nfl standings top](#)

nfl standings review

[netflix guide returns](#)

[student loan repayment guide](#)

[fantasy football update tutorial](#)

remote jobs how to

top movies near me

walking workout prices open now

prime big deal days price

goodreads choice top

[viral cozy mystery same day delivery](#)

An Investigation Of Web Crawler Behavior Characterization :

La regola dell'equilibrio Award-winning, best-selling novelist Gianrico Carofiglio was born in Bari in 1961 and worked for many years as a prosecutor specializing in organized crime. La regola dell'equilibrio by Carofiglio, Gianrico Carofiglio is best known for the Guido Guerrieri crime series: Involuntary Witness, A Walk in the Dark, Reasonable Doubts, Temporary Perfections and now, A Fine ... La regola dell'equilibrio La regola dell'equilibrio è un romanzo giallo scritto da Gianrico Carofiglio e pubblicato da Einaudi nel 2014. Fa parte della serie I casi dell'avvocato ... La regola dell'equilibrio Quotes by Gianrico Carofiglio The man who lies to himself and listens to his own lie comes to such a pass that he can no longer distinguish the truth, within him or around him. La regola dell'equilibrio book by Gianrico Carofiglio "A FINE LINE is a terrific novel, a legal thriller that is also full of complex meditations on the life of the lawyer and the difficult compromises inherent ... La regola dell'equilibrio - Hardcover La regola dell'equilibrio ISBN 13: 9788806218126. La regola dell'equilibrio - Hardcover. 3.84 avg rating • (1,891 ratings by Goodreads). View all 41 copies ... La regola dell'equilibrio by Gianluca Carofiglio: Good Used book that is in clean, average condition without any missing pages. Seller Inventory # 46077381-6. Contact seller · Report this item. La regola dell'equilibrio Dec 11, 2014 — Guido Guerrieri is a changed man. Handed a troubling medical diagnosis, his usual confidence has been shaken, and even if the jury is out on ... La regola dell'equilibrio by Gianrico Carofiglio | eBook eBook(Italian-language Edition) · \$8.99. La regola dell'equilibrio, Brand New, Free shipping in the US La regola dell'equilibrio, Brand New, Free shipping in the US · Great Book Prices Store (274250) · 97.3% positive feedback ... Química. Solucionario. Chang & Goldsby. 11va edición. ... (Chemistry. Solutions manual. 11th edition). 697

Pages. Química. Solucionario. Chang & Goldsby. 11va edición. (Chemistry. Solutions manual. 11th edition) ... Chemistry - 11th Edition - Solutions and Answers Find step-by-step solutions and answers to Chemistry - 9780073402680, as well as thousands of textbooks so you can move forward with confidence. Student Solutions Manual for Chemistry by Raymond ... Student Solutions Manual for Chemistry by Raymond Chang (2012-01-19) [Raymond Chang; Kenneth Goldsby;] on Amazon.com. *FREE* shipping on qualifying offers. Student Solutions Manual for Chemistry by Chang, Raymond The Student Solutions Manual is written by Brandon J. Cruickshank (Northern Arizona University), Raymond Chang, and Ken Goldsby. Student solutions manual to accompany Chemistry ... Student solutions manual to accompany Chemistry, eleventh edition, [by] Raymond Chang, Kenneth A. Goldsby. Show more ; Genre: Problems and exercises ; Physical ... Student Solutions Manual for Chemistry | Rent Student Solutions Manual for Chemistry 11th edition ; ISBN-13: 9780077386542 ; Authors: Raymond Chang, Kenneth Goldsby ; Full Title: Student Solutions Manual for ... Student Solutions Manual For Chemistry 11th Edition ... Access Student Solutions Manual for Chemistry 11th Edition Chapter 10 Problem 95P solution now. Our solutions are written by Chegg experts so you can be ... Chemistry - Student Solution Manual 11th edition The Student Solutions Manual is written by Brandon J. Cruickshank (Northern Arizona University), Raymond Chang, and Ken Goldsby. Raymond Goldsby Chang | Get Textbooks Student Solutions Manual for Chemistry (11th Edition) by Raymond Chang, Kenneth A. Goldsby, Brandon Cruickshank, Robert Powell Paperback, 656 Pages ... solutions-manual-chemistry-chapter-11 Chemistry Chang 11th Edition Solutions Manual Click here to download the 11th ISBN-10: 0073402680 Type: Solutions Manual This is a sample chapter. 11. Financial Accounting Theory by Scott, William William Scott. Financial Accounting Theory. 7th Edition. ISBN-13: 978-0132984669, ISBN-10: 0132984660. 4.7 4.7 out of 5 stars 47 Reviews. 3.6 on Goodreads. (65). William R. Scott | FINANCIAL ACCOUNTING THEORY Financial accounting theory / William R. Scott. - Seventh edition. Includes bibliographical references and index. ISBN 978-0-13-298466-9 (bound). Financial Accounting Theory (7th... by William Rufus Scott Financial Accounting Theory (7th Edition) by William R. Scott (2015-02-20) ; Payment. Secure transaction ; Print length. 0 pages ; Publisher. Pearson ; Publication ... Financial Accounting Theory - Scott, William Financial Accounting Theory provides a thorough presentation of financial accounting theories. This new edition continues to include considerable coverage ... Results for "Scott Financial-Accounting-Theory-7th-Edition" Search results. Financial Accounting Theory. 8th Edition. William R. Scott, Patricia O'Brien. ISBN-13: 9780134166681. Print for £187.56. Search results. We didn't ... Financial Accounting Theory | Rent | 9780132984669 ISBN-13: 9780132984669 ; Authors: William R Scott, William Scott ; Full Title: Financial Accounting Theory ; Edition: 7th edition ; ISBN-13: 978-0132984669. Financial accounting theory | WorldCat.org Financial accounting theory ; Author: William R. Scott ; Edition: 7. ed View all formats and editions ; Publisher: Pearson, Toronto, 2015. Financial Accounting Theory (7th Edition) (Hardcover) Financial Accounting Theory (7th Edition) (Hardcover); Author: by William R. Scott; Book Condition: Used - Fine; Quantity Available: 1; Edition: 7th; Binding ...

Financial Accounting Theory by William R. Scott This newly revised text provides a theoretical approach to financial accounting in Canada, without overlooking institutional structure and standard setting. Financial Accounting Theory (7th Edition) - AbeBooks Synopsis: Financial Accounting Theory provides a thorough presentation of financial accounting theories. This new edition continues to include considerable ...