

Simple Algorithm for Page Layout Analysis

A. O. Shigarov and R. K. Fedorov

*Institute for System Dynamics and Control Theory, Siberian Branch, Russian Academy of Sciences,
ul. Lermontova 134, Irkutsk, 664033 Russia*

Abstract—An algorithm for page layout analysis (segmentation) is suggested in the paper. It allows whitespace between text blocks to be detected on a document page. The algorithm could be used in document analysis and recognition problems. In particular, it can be used for column recognition in multicolumn text and tables. The suggested algorithm is quite simple for implementation.

INTRODUCTION

The number of documents in the world is growing rapidly; this is discussed in [6] in particular. Document analysis and recognition (DAR) systems are developed for automation of information extraction from document images [7]. Document layout analysis or page segmentation is used for dividing document into specific parts (e.g., columns, figures, tables). This is an important DAR problem. Different methods of document layout analysis are discussed in [3].

Conventionally, there are two approaches to the problem of page or table column segmentation. The first approach is to analyze of text layout (text blocks); this usually requires the use of complex data structures. For example, such an approach with the use of a Voronoi diagram for page segmentation is suggested in [5]. The second approach is to analyze whitespace (free of text) on a page. Gaps separate text and table columns on a page, as is shown in Fig. 1. Algorithms using whitespace analysis are suggested in [1, 2, 8]. Algorithms [1, 8] are briefly considered in [2]. They can be used for detecting whitespace between text blocks. The authors of [2] point out that the algorithms [1, 8] are difficult for implementation. A geometrical layout analysis algorithm, simple for implementation, is presented in [2]; it provides for detecting whitespace gaps in multicolumn text and is described in terms of the largest empty rectangle problem [4]. The algorithm input is a bounding box including obstacles (rectangles). The algorithm [2] finds the largest empty rectangle among obstacles inside the bounding box. To find whitespace gaps on a page, n -better solutions (empty rectangles) in descending order are found on the page. The algorithm [2] is greedy. The found i -largest empty rectangle becomes an obstacle when searching for the next $i + 1$ -largest empty rectangle. However, it is probable that some of n -better solutions (largest rectangles found) are not gaps between columns, but, e.g., horizontal gaps between paragraphs or

tables on the page. This is the disadvantage of the algorithm [2].

In the paper, we suggest an unconventional simple algorithm for detecting whitespace gaps on a document page. It allows detecting vertical gaps (visually they are stretched up vertically), as well as horizontal gaps if the X and Y axis are exchanged. The suggested algorithm is simple; its Object Pascal implementation consists of about 60 lines of code (expressions).

1. PROBLEM FORMULATION

The geometrical objects considered in this work are presented in the Cartesian coordinate system, where the x -coordinates increase from left to right and y -coordinates increase from the top down. The following assumptions are used in this work. The rectangle (e.g., an obstacle, bounding box, or gap) $r = (x_1, y_1, x_2, y_2)$ is defined by the coordinates of its sides (boundaries): left $x_1 = x_1(r)$, top $y_1 = y_1(r)$, right $x_2 = x_2(r)$, and bottom $y_2 = y_2(r)$; in addition, its sides are parallel to the corresponding coordinate axes. The vertical line $l = (x, y_1, y_2)$ is normal to the X axis and is defined by its coordinates: $x = x(l)$ by the x -coordinate, $y_1 = y_1(l)$ by the minimum (top) y -coordinate, and $y_2 = y_2(l)$ by the maximum (bottom) y -coordinate.

It is assumed that the bounding rectangle b and the set of obstacles (rectangles) $R = \{r_1, \dots, r_n\}$, $n \in \mathbb{N}$ are specified. The rectangle b usually bounds a document page or its part (e.g., a table), while the obstacles are bounding rectangles for text blocks (e.g., words or lines). The obstacles from the set R are totally inside the bounding rectangle b and do not overlap each other.

Let us define two sets of obstacles (rectangles) $R^+ = \{r_1, r_2, \dots, r_m, r_3\}$ and $R^- = \{r_1, r_2, \dots, r_m, r_3\}$, where $r_1 = (x_1, y_1, x_2, y_2)$, $r_2 = (x_m, y_m, x_m, y_2)$, $r_3 = (x_2, y_1, x_2, y_2)$, $x_1 = x_1(b)$, $y_1 = y_1(b)$, $x_2 = x_2(b)$, and $y_2 = y_2(b)$. Let us also define the gap as a rectangle bounding a certain part of whitespace inside the bounding rectangle b .

Simple Algorithm Page Layout Analysis Springerlink

Hongru Du



Simple Algorithm Page Layout Analysis Springerlink:

Decoding **Simple Algorithm Page Layout Analysis Springerlink**: Revealing the Captivating Potential of Verbal Expression

In a time characterized by interconnectedness and an insatiable thirst for knowledge, the captivating potential of verbal expression has emerged as a formidable force. Its capability to evoke sentiments, stimulate introspection, and incite profound transformations is genuinely awe-inspiring. Within the pages of "**Simple Algorithm Page Layout Analysis Springerlink**," a mesmerizing literary creation penned by a celebrated wordsmith, readers set about an enlightening odyssey, unraveling the intricate significance of language and its enduring affect our lives. In this appraisal, we shall explore the book is central themes, evaluate its distinctive writing style, and gauge its pervasive influence on the hearts and minds of its readership.

https://wwwnew.greenfirefarms.com/files/uploaded-files/Download_PDFS/Quick%20Us%20National%20Parks%20Guide%20For%20Workers%2015101.pdf

Table of Contents Simple Algorithm Page Layout Analysis Springerlink

1. Understanding the eBook Simple Algorithm Page Layout Analysis Springerlink
 - The Rise of Digital Reading Simple Algorithm Page Layout Analysis Springerlink
 - Advantages of eBooks Over Traditional Books
2. Identifying Simple Algorithm Page Layout Analysis Springerlink
 - 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 a Simple Algorithm Page Layout Analysis Springerlink
 - User-Friendly Interface
4. Exploring eBook Recommendations from Simple Algorithm Page Layout Analysis Springerlink
 - Personalized Recommendations

- Simple Algorithm Page Layout Analysis Springerlink User Reviews and Ratings
- Simple Algorithm Page Layout Analysis Springerlink and Bestseller Lists
- 5. Accessing Simple Algorithm Page Layout Analysis Springerlink Free and Paid eBooks
 - Simple Algorithm Page Layout Analysis Springerlink Public Domain eBooks
 - Simple Algorithm Page Layout Analysis Springerlink eBook Subscription Services
 - Simple Algorithm Page Layout Analysis Springerlink Budget-Friendly Options
- 6. Navigating Simple Algorithm Page Layout Analysis Springerlink eBook Formats
 - ePub, PDF, MOBI, and More
 - Simple Algorithm Page Layout Analysis Springerlink Compatibility with Devices
 - Simple Algorithm Page Layout Analysis Springerlink Enhanced eBook Features
- 7. Enhancing Your Reading Experience
 - Adjustable Fonts and Text Sizes of Simple Algorithm Page Layout Analysis Springerlink
 - Highlighting and Note-Taking Simple Algorithm Page Layout Analysis Springerlink
 - Interactive Elements Simple Algorithm Page Layout Analysis Springerlink
- 8. Staying Engaged with Simple Algorithm Page Layout Analysis Springerlink
 - Joining Online Reading Communities
 - Participating in Virtual Book Clubs
 - Following Authors and Publishers Simple Algorithm Page Layout Analysis Springerlink
- 9. Balancing eBooks and Physical Books Simple Algorithm Page Layout Analysis Springerlink
 - Benefits of a Digital Library
 - Creating a Diverse Reading Collection Simple Algorithm Page Layout Analysis Springerlink
- 10. Overcoming Reading Challenges
 - Dealing with Digital Eye Strain
 - Minimizing Distractions
 - Managing Screen Time
- 11. Cultivating a Reading Routine Simple Algorithm Page Layout Analysis Springerlink
 - Setting Reading Goals Simple Algorithm Page Layout Analysis Springerlink
 - Carving Out Dedicated Reading Time
- 12. Sourcing Reliable Information of Simple Algorithm Page Layout Analysis Springerlink
 - Fact-Checking eBook Content of Simple Algorithm Page Layout Analysis Springerlink

- 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

Simple Algorithm Page Layout Analysis Springerlink Introduction

In this digital age, the convenience of accessing information at our fingertips has become a necessity. Whether its research papers, eBooks, or user manuals, PDF files have become the preferred format for sharing and reading documents. However, the cost associated with purchasing PDF files can sometimes be a barrier for many individuals and organizations. Thankfully, there are numerous websites and platforms that allow users to download free PDF files legally. In this article, we will explore some of the best platforms to download free PDFs. One of the most popular platforms to download free PDF files is Project Gutenberg. This online library offers over 60,000 free eBooks that are in the public domain. From classic literature to historical documents, Project Gutenberg provides a wide range of PDF files that can be downloaded and enjoyed on various devices. The website is user-friendly and allows users to search for specific titles or browse through different categories. Another reliable platform for downloading Simple Algorithm Page Layout Analysis Springerlink free PDF files is Open Library. With its vast collection of over 1 million eBooks, Open Library has something for every reader. The website offers a seamless experience by providing options to borrow or download PDF files. Users simply need to create a free account to access this treasure trove of knowledge. Open Library also allows users to contribute by uploading and sharing their own PDF files, making it a collaborative platform for book enthusiasts. For those interested in academic resources, there are websites dedicated to providing free PDFs of research papers and scientific articles. One such website is Academia.edu, which allows researchers and scholars to share their work with a global audience. Users can download PDF files of research papers, theses, and dissertations covering a wide range of subjects. Academia.edu also provides a platform for discussions and networking within the academic community. When it comes to downloading Simple Algorithm Page Layout Analysis Springerlink free PDF files of magazines, brochures, and catalogs, Issuu is a popular choice. This digital publishing platform hosts a vast collection of publications from around the world. Users can search for specific titles or explore various categories and genres. Issuu offers a seamless reading experience with its user-friendly interface and allows users to download PDF files for offline reading. Apart from dedicated platforms, search engines also play a crucial role in finding free

PDF files. Google, for instance, has an advanced search feature that allows users to filter results by file type. By specifying the file type as "PDF," users can find websites that offer free PDF downloads on a specific topic. While downloading Simple Algorithm Page Layout Analysis Springerlink free PDF files is convenient, it's important to note that copyright laws must be respected. Always ensure that the PDF files you download are legally available for free. Many authors and publishers voluntarily provide free PDF versions of their work, but it's essential to be cautious and verify the authenticity of the source before downloading Simple Algorithm Page Layout Analysis Springerlink. In conclusion, the internet offers numerous platforms and websites that allow users to download free PDF files legally. Whether it's classic literature, research papers, or magazines, there is something for everyone. The platforms mentioned in this article, such as Project Gutenberg, Open Library, Academia.edu, and Issuu, provide access to a vast collection of PDF files. However, users should always be cautious and verify the legality of the source before downloading Simple Algorithm Page Layout Analysis Springerlink any PDF files. With these platforms, the world of PDF downloads is just a click away.

FAQs About Simple Algorithm Page Layout Analysis Springerlink 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. Simple Algorithm Page Layout Analysis Springerlink is one of the best book in our library for free trial. We provide copy of Simple Algorithm Page Layout Analysis Springerlink in digital format, so the resources that you find are reliable. There are also many Ebooks of related with Simple Algorithm Page Layout Analysis Springerlink. Where to download Simple Algorithm Page Layout Analysis Springerlink online for free? Are you looking for Simple Algorithm Page Layout Analysis Springerlink PDF? This is definitely going to save you time and cash in something you should think about.

Find Simple Algorithm Page Layout Analysis Springerlink :

quick us national parks guide for workers 15101

affordable ai video generator 2025 for workers 14506

advanced pilates for beginners 2025 for students 15517

why us national parks guide for experts 15556

how to start home workout 2025 for beginners 14109

pro ai tools for small business for beginners 14225

easy ai seo tools guide for beginners 15431

top method for side hustles for students for creators 15065

top method for capsule wardrobe for small business for experts 15460

expert ai seo tools explained for creators 13718

quick cheap flights usa ideas for workers 15060

advanced sleep hygiene tips online for students 14852

pro ai video generator for small business for experts 15003

~~how to use ai seo tools explained for creators 13836~~

top method for home workout step plan for students 15520

Simple Algorithm Page Layout Analysis Springerlink :

Financial Reporting, Financial Statement Analysis And ... Access Financial Reporting, Financial Statement Analysis and Valuation 7th Edition solutions now. Our solutions are written by Chegg experts so you can be ... Solution Manual for Financial Reporting ... - Course Hero View Solution Manual for Financial Reporting, Financial Statement Analysis and Valuation A Strategic Pers from ECONO 221 at Universit  di Roma Tor Vergata. Financial Reporting and Analysis 7th Edition Revsine ... Full download : <http://goo.gl/s7uYSK> Financial Reporting and Analysis 7th Edition Revsine Solutions Manual, 7th Edition, Collins, Financial Reporting and ... Financial Reporting Financial Statement Analysis and ... Apr 10, 2019 — Financial Reporting Financial Statement Analysis and Valuation 7th Edition Whalen Solutions Manual Full Download: <http://alibabadownload.com> ... Solution Manual for Financial Reporting and Analysis 7th ... Solution Manual For Financial Reporting and Analysis 7th Edition by Revsine ... uses of financial statement information (e.g., valuation, credit analysis, and solutions manual, test bank for Financial Reporting ... solutions manual, test bank for Financial Reporting, Financial Statement Analysis and Valuation A Strategic Perspective 7e 7/E 7th edition by James Wahlen ... Solution Manual for

Financial Reporting Solution Manual for Financial Reporting Financial Statement Analysis and Valuation 9th Edition by Wahlen - Free download as PDF File (.pdf), ... Epub free Financial reporting statement analysis and ... Apr 10, 2023 — analysis and valuation solution manual. (2023). Business Analysis & Valuation Business Analysis and Evaluation Functional Analysis and. Financial Reporting and Analysis 7th Edi - 2 Financial Analysis financial reporting and analysis 7th edition revsine solutions manual full download: financial. Solution Manual Financial Reporting ... Aug 30, 2018 — Solution Manual Financial Reporting Financial Statement Analysis and Valuation 7th Edition by James M. Whalen. Solution Manual. Theatre: Brief Version, 10th Edition - Amazon.com Robert Cohen's Theatre Brief, 10th Edition continues to provide an insiders guide to the world of theatre, where students are given a front-row seat. This ... Theatre, 10th Edition - Cohen, Robert: Books Robert Cohen's Theatre, 10th Edition continues to provide an insider's guide to the world of theatre, where students are given a front-row seat. Theatre, 10th Edition - Cohen, Robert - AbeBooks Robert Cohen's Theatre, 10th Edition continues to provide an insider's guide to the world of theatre, where students are given a front-row seat. theatre 10th edition Theatre, 10th Edition by Cohen, Robert and a great selection of related books, art and collectibles available now at AbeBooks.com. Theatre: Brief Version 10th Edition By Robert Cohen Theatre: Brief Version 10th Edition By Robert Cohen. Theatre: Brief Version, 10th Edition - Paperback, by Cohen ... Theatre: Brief Version, 10th Edition - Paperback, by Cohen Robert - Good ; Book Title. Theatre: Brief Version, 10th Edition ; ISBN. 9780077494261 ; Publication ... Theatre: Brief Version, 10th Edition by Cohen, Robert ... From the publisher ... Robert Cohen's Theatre Brief, 10th Edition continues to provide an insiders guide to the world of theatre, where students are given a front ... Theatre 10th Edition Robert Cohen What I Thought I Knew. Woman and Scarecrow. The Creation of the Mods and Rockers. Theatre, Brief Loose Leaf. Reflections on Berkeley in the 1960s. Theatre, Brief Edition - ROBERT COHEN Apr 20, 2023 — Tenth Edition McGraw-Hill, 2013. A condensation of the full version of Cohen's best-selling Theatre, which includes all of its chapters on ... 9780073514222 - Theatre Loose Leaf by Robert Cohen Robert Cohen's Theatre, 10th Edition continues to provide an insider's guide to the world of theatre, where students are given a front-row seat. This lively ... The Unfinished Revolution: Philip Gould In May 1997, the Conservatives were ejected from British office after 18 years in power, and the Labour Party which replaced them had itself changed ... How the Modernisers Saved the Labour Party by Philip Gould Written in 1999, the title has become ironic over time - Blair's revolution remained unfinished, as the promise of his program was compromised by Iraq and ... The Unfinished Revolution: How New Labour... by Gould ... In May 1997, the Conservatives were ejected from British office after 18 years in power, and the Labour Party which replaced them had itself changed irrevocably ... How the Modernisers Saved the Labour Party - Philip Gould The Unfinished Revolution: How the Modernisers Saved the Labour Party ... On 1 May 1997, an event regarded by many as a sea-change occurred in British politics. The Unfinished Revolution: How the Modernisers Saved ... The Unfinished Revolution: How the Modernisers Saved the Labour Party - Softcover ; Publisher: Little, Brown Book Group, 1999 ; Buy Used

Condition: Very Good How the Modernisers Saved the Labour Party - Philip Gould Philip Gould is a political strategist and polling adviser who has worked with the Labour leadership since the 1980s. In this book he describes its rise and ... The Unfinished Revolution by Philip Gould The Unfinished Revolution is the definitive story of New Labour from its genesis to its election defeat 2010 - covering over 25 years and six general ... how the modernisers saved the Labour Party / Philip Gould ... The unfinished revolution : how the modernisers saved the Labour Party / Philip Gould. Request Order a copy. Bib ID: 2206389; Format: Book; Author: Gould ... The unfinished revolution : how the modernisers saved ... Nov 27, 2020 — The unfinished revolution : how the modernisers saved the Labour Party. by: Gould, Philip, 1950-2011. Publication date: 1999. Topics: Labour ... How the Modernisers Sav... by Gould, Philip Paperback ... The Unfinished Revolution: How the Modernisers Saved the Labour Party. Book Binding:Paperback / softback. Author:Gould, Philip. We all like the idea of saving ...