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) $r = (x_l, y_t, x_r, y_b)$ is defined by the coordinates of its sides (boundaries): left $c_l = x_l(r)$, top $y_t = y_t(r)$, right $x_r = x_r(r)$, and bottom $y_b = y_b(r)$; in addition, its sides are parallel to the corresponding coordinate axes. The vertical line $l = (x, y_t, y_b)$ is normal to the X axis and is defined by its coordinates: $x = x(l)$ by the x-coordinate, $y_t = y_t(l)$ by the minimum (top) y-coordinate, and $y_b = y_b(l)$ by the maximum (bottom) y-coordinate.

It is assumed that the bounding rectangle b and the set of obstacles (rectangles) $R = \{r_1, ..., r_n\}$, $n \in \mathbb{N}$ 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) $R = \{r_0, r_1, ..., r_n, r_3\}$ and $\mathbf{R}^m = \{r_0, r_1, ..., r_n, r_3\}$, where $r_0 = (x_l, y_t, x_r, y_b)$, $r_1 = (x_l, y_t, x_r, y_b)$, $r_2 = (x_l, y_t, x_r, y_b)$, $r_3 = (x_l, y_t, x_r, y_b)$, $x_l = x_l(b)$, $y_t = y_t(b)$, $x_r = x_r(b)$, and $y_b = y_b(b)$. 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](#)

Helmut Knebl
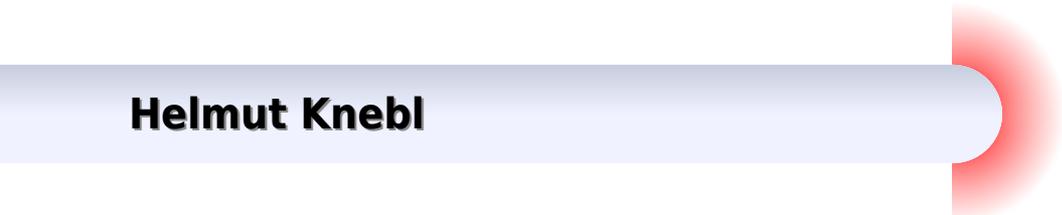
Simple Algorithm Page Layout Analysis Springerlink:

Simple Algorithm Page Layout Analysis Springerlink Book Review: Unveiling the Magic of Language

In an electronic digital era where connections and knowledge reign supreme, the enchanting power of language has be much more apparent than ever. Its capability to stir emotions, provoke thought, and instigate transformation is really remarkable. This extraordinary book, aptly titled "**Simple Algorithm Page Layout Analysis Springerlink**," written by a very acclaimed author, immerses readers in a captivating exploration of the significance of language and its profound effect on our existence. Throughout this critique, we shall delve into the book is central themes, evaluate its unique writing style, and assess its overall influence on its readership.

https://socketapi.adit.com/About/book-search/HomePages/biology%20laboratory%20a%20chapter%2014%20making%20karyotypes%20answers.pdf

Table of Contents Simple Algorithm Page Layout Analysis Springerlink

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, its 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 its 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 its 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

What is a Simple Algorithm Page Layout Analysis Springerlink PDF? A PDF (Portable Document Format) is a file format developed by Adobe that preserves the layout and formatting of a document, regardless of the software, hardware, or operating system used to view or print it. **How do I create a Simple Algorithm Page Layout Analysis Springerlink PDF?** There are several ways to create a PDF: Use software like Adobe Acrobat, Microsoft Word, or Google Docs, which often have built-in PDF creation tools. Print to PDF: Many applications and operating systems have a "Print to PDF" option that allows you to save a document as a PDF file instead of printing it on paper. Online converters: There are various online tools that can convert different file types to PDF. **How do I edit a Simple Algorithm Page Layout Analysis Springerlink PDF?** Editing a PDF can be done with software like Adobe Acrobat, which allows direct editing of text, images, and other elements within the PDF. Some free tools, like PDFescape or Smallpdf, also offer basic editing capabilities. **How do I convert a Simple Algorithm Page Layout Analysis Springerlink PDF to another file format?** There are multiple ways to convert a PDF to another format: Use online converters like Smallpdf, Zamzar, or Adobe Acrobats export feature to convert PDFs to formats like Word, Excel, JPEG, etc. Software like Adobe Acrobat, Microsoft Word, or other PDF editors may have options to export or save PDFs in different formats. **How do I password-protect a Simple Algorithm Page Layout Analysis Springerlink PDF?** Most PDF editing software allows you to add password protection. In Adobe Acrobat, for instance, you can go to "File" -> "Properties" -> "Security" to set a password to restrict access or editing capabilities. Are there any free alternatives to Adobe Acrobat for working with PDFs? Yes, there are many free alternatives for working with PDFs, such as: LibreOffice: Offers PDF editing features. PDFsam: Allows splitting, merging, and editing PDFs. Foxit Reader: Provides basic

PDF viewing and editing capabilities. How do I compress a PDF file? You can use online tools like Smallpdf, ILovePDF, or desktop software like Adobe Acrobat to compress PDF files without significant quality loss. Compression reduces the file size, making it easier to share and download. Can I fill out forms in a PDF file? Yes, most PDF viewers/editors like Adobe Acrobat, Preview (on Mac), or various online tools allow you to fill out forms in PDF files by selecting text fields and entering information. Are there any restrictions when working with PDFs? Some PDFs might have restrictions set by their creator, such as password protection, editing restrictions, or print restrictions. Breaking these restrictions might require specific software or tools, which may or may not be legal depending on the circumstances and local laws.

Find Simple Algorithm Page Layout Analysis Springerlink :

biology laboratory a chapter 14 making karyotypes answers
books bangla stories
book poverty and development into the 21st century pdf
bom dia 365 mensagens
biological physics nelson solutions pdf
bodybuilding supplements yes or no bodybuilding supplements guide for men and women pre and post workout steroids and more
bitcoin beginner secrets the simple step by step to making money with bitcoins
biology o level question answers
bmw 3 series e90 e91 e92 e93 service manual 2006 2010 free
binary options strategies for directional and volatility trading
black decker the complete guide to concrete masonry 4th edition build with concrete brick block natural stone
black decker complete guide
biological effects of electric and magnetic fields
blockchain link springer
biochar for environmental management science technology and implementation
bogie designs skf

Simple Algorithm Page Layout Analysis Springerlink :

Magnets and Motors Teacher's Guide Magnets and Motors Teacher's Guide ... Only 1 left in stock - order soon. ... Shows a

little shelf wear. Cover, edges, and corners show the most. Pages are clean ... Magnets and Motors: Teacher's Guide A powerful way to foster appreciation for the impact of science and critical and innovative thinking is through art and the humanities. Learn more about the ... Magnets and Motors: Teacher's Guide Jan 1, 1991 — Magnets and Motors: Teacher's Guide · From inside the book · Contents · Common terms and phrases · Bibliographic information. Title ... Magnets and Motors Teacher's Guide - National Science ... Magnets and Motors Teacher's Guide by National Science Resources Center - ISBN 10: 0892786922 - ISBN 13: 9780892786923 - National Academy of Sciences. STC Assessment Guide: Magnets and Motors Daily formative assessments gauge student knowledge and let you know whether they are grasping key science concepts. The 15-to 20-question summative assessment ... STC MAGNETS & MOTORS KIT Mar 30, 2015 — Magnets & Motors – 6th Grade. NGSS Curriculum Redesign. 6th magnets and motors - UNIT GUIDE. 46. 3/30/2015 11:40 PM. Science of Electricity ... Magnet Motors Teacher Guide – Green Design Lab Magnet Motors Teacher Guide · Related Articles · Our Programs. Magnets and Electricity STEM, Free PDF Download Our Magnets and Electricity STEM lesson plan explores the world of electromagnetism and teaches students how this phenomenon works. Free PDF download! Lesson By Lesson Guide Magnetism & Electricity (FOSS Kit) It is helpful to model connections with the D-Cell and motor for students. ... Teachers Guide. Science Notebook Helper. - Students record the focus question ... 10-Easy-Steps-to-Teaching-Magnets-and-Electricity.pdf Mar 19, 2020 — Electric Motors. Objective: To learn how an electric motor works by building one. In addition to the great lessons and experiments, this book ... Manual of Neonatal Care (7th Edition) by JP Cloherty · Cited by 919 — Materials appearing in this book prepared by individuals as part of their official duties as U.S. government employees are not covered by the ... Manual of neonatal care : Free Download, Borrow, and ... Oct 16, 2021 — xxii, 1007 p. : 21 cm "This edition of the Manual of Neonatal Care has been completely updated and extensively revised to reflect the ... A Manual of Neonatal Intensive Care The information or guidance contained in this book is intended for use by medical, scientific or health-care professionals and is provided strictly as a ... NEONATAL CARE CLINICAL GUIDELINES This first edition of our national neonatal care clinical guidelines is an initiative that aims to ensure that all the neonates in the Kingdom of Eswatini are ... NEONATAL MANUAL FOR STANDARD NEWBORN CARE This Operations Manual was produced by the INTERGROWTH-21st Neonatal Group, based on the 1st Meeting of the Neonatal Group, Oxford, July 2009. Manual of neonatal care : Free Download, Borrow, and ... Oct 13, 2020 — Manual of neonatal care · Share or Embed This Item · Flag this item for · Manual of neonatal care · DOWNLOAD OPTIONS · IN COLLECTIONS · SIMILAR ... Care of the Newborn Reference Manual by D Beck · 2004 · Cited by 9 — SAVING NEWBORN LIVES is a 10-15 year global initiative of. Save the Children to improve the health and survival of newborns in the developing world. Ovid - Cloherty and Stark's Manual of Neonatal Care Practical, informative, and easy to read, Cloherty and Stark's Manual of Neonatal Care , 9th Edition, offers an up-to-date approach to the diagnosis and ... Neonatal Clinical Practice Guidelines 2018-2021 Original These guidelines have

been developed, at the request of the Ministry of Health, as an aide- memoire for all staff concerned with the management of neonates to ... NICU Portal: Selected eBooks - Darnall Medical Library Dec 4, 2023 — Can I download or print an eBook? It depends on the company providing ... Cloherty and Stark's Manual of Neonatal Care. Pdms 2 scoring manual Peabody developmental motor scales and activity cards. Pdms standard scores. Pdms 2 scoring manual pdf. Publication date: 2000 Age range: Birth through age 5 ... Guidelines to PDMS-2 Raw Scores: • Add scores from each subtest evaluated. –Example Grasping and Visual-Motor are subtests for fine motor evaluations. Peabody Developmental Motor Scales, Third Edition The PDMS-3 norms are based on an all-new sample of ... There are no tables in the PDMS-3 manual – all scores are calculated using the online scoring system. (PDMS-2) Peabody Developmental Motor Scales, Second ... Benefit. Assesses both qualitative and quantitative aspects of gross and fine motor development in young children; recommends specific interventions ; Norms. Peabody Developmental Motor Scales–Third Edition ... The PDMS-3 Online Scoring and Report System yields four types of normative scores: ... The PDMS-3 norms are based on an all-new sample of 1,452 children who were ... Peabody Developmental Motor Scale (PDMS-2) This subtest measures a child's ability to manipulate balls, such as catching, throwing and kicking · These skills are not apparent until a child is 11 months ... PDMS-2 Peabody Developmental Motor Scales 2nd Edition Access three composite scores: Gross Motor Quotient, Fine Motor Quotient, and Total Motor Quotient. Helps facilitate the child's development in specific skill ... PDMS-2 Peabody Developmental Motor Scales 2nd Edition Norms: Standard Scores, Percentile Ranks, and Age ... Access three composite scores: Gross Motor Quotient, Fine Motor Quotient, and Total Motor Quotient. Peabody Developmental Motor Scales High scores on this composite are made by children with well-developed gross motor abilities. These children would have above average movement and balance ...