An Efficient Approach for E-Content Management and Delivery in Digital Library

Achintya K Mandal
Deputy Registrar, Assam University, Silchar - 788011, INDIA
Email: achintya99@hotmail.com

Utpal Roy
Department of Computer and System Sciences, Visva-Bharati University, Santiniketan, INDIA,
Email: utpal.roy@visva-bharati.ac.in

World Digital Libraries 5(1): 1–10

Abstract

In a digital library environment, the delivery and management of e-content is still a challenging task. The present study addresses an easy-to-implement technique applicable for the management of e-content and its delivery in digital library. The experiment has been performed upon some paper-based derived digital documents of many variants. The idea behind it is that the non-text region of a book-page, mainly the white margin around four sides and the space between two paragraphs have been discarded during scanning and then the conventional algorithm has been applied for removing the obvious noise incurred during scanning process. In essence, a reasonable amount of storage space is saved when such a document is placed in the secondary memory of the digital library server. Obviously, the web request for such pages is served in a better speed than usual. Moreover, a margin resetting algorithm in client side is applied to return these pages in its original form. Considering, mainly the storage space and the access speed while serving web request, the obtained results have some positive implications towards the problem of e-content management and its delivery in digital library system.

Keywords: Digital Library, Content Management, Content Delivery, Derived Digital Document and Document Images.

1. Introduction

The digital library today uses wealth of different document formats storing and representing the content. The genesis and development of digital library is a long saga and has presently become a very common source for information seekers. Digital libraries come in many forms. They may contain text of the document instead of only simple metadata or catalogues of bibliographical information. They can also contain images, audio, or multimedia materials. All this information may be available in different formats, created with different softwares. The resources may reside on different servers using no unified thesauri or heterogeneous indexing schemes. All this makes information retrieval a very complex process. Every information system is unique when it comes to retrieval methods, and it is more or less necessary to have a fair idea of the characteristic features of each system to be able to perform a relevant search. This becomes even more complex when some digital libraries allow users to conduct search across a range of distributed services as described by Chowdhury and Chowdhury (2003). With many general issues of digital object storage, moderately large and secured storage system is an important and challenging aspect of modern day digital libraries. Storing of digital objects leads to a large byte-count. An institution's digital repository (which may have a label such as 'library' or 'archive') may include terabytes of data spread across thousands of objects in hundreds of formats. It may, for ease of handling and control, transfer the contents to a more readily managed medium. The sheer ease of producing and proliferating electronic files, combined with the perception that storage is so cheap, leads to an uncontrolled explosion in numbers and storage volumes, erratic and unplanned use of storage devices. The 'Total Cost of Ownership' (TCO) of storage (which includes the life-long cost of its management) varies from organization to organization, but can only be minimized if the techniques for minimizing the size of the content are developed. Content management in digital library has been discussed by many researchers. Wu and Liu (2001) discussed Internet-based e-content management. Specifically, they reviewed the technologies, the criteria, the issues and concerns in content repository, content contribution, workflow, automation services, and lifecycle of automation services for controlling managing content and processes. Warren and Alsmeyer (2005) describe the application of semantic knowledge technology to a case study in intelligent content management, specifically the BT digital library.

In this context, the *Trove (launched in 2009)* is an important discovery by Australians for their libraries. It harvests metadata from over 1,000 Australian libraries and other cultural heritage organizations, allowing free public access to over 100 million items. The guiding principle of Trove is 'Find and Get' (Holley, R, 2010). The details of its application and development has been described by Holley, Rose (Holley, Rose 2011)

Most of the libraries have a good collection of old and rare books, as there is no electronic equivalent of those collections. The only way to integrate them into digital library is by scanning the individual pages (document images) with a good quality scanner preferably by a planetary scanner. The file size of the document images is very large in comparison to the electronic documents, though it depends on the resolution of the scanning and the image file format chosen for saving. In normal practice, the document images are not properly processed before sending for final use of the service. Optical Character Recognition (OCR) software plays an important role here to convert the document images into its electronic equivalent. As a result the file size reduces to a great extent and all the words of the document become electronically searchable. Several OCR software for European languages particularly in English Nagy (2000), is available, that can also handle multi-columns, images, tables, and also preserve the document layout analysis. On the other hand, they work

on document analysis and recognition in other languages particularly Indian languages are still in a nascent form. However, a few basic OCR modules are available. Most of them, from north Indian Brahmi based scripts – such as Devnagari (Hindi), Bengali (Bangla) as discussed by Chaudhuri, Garain and Mitra. Ray Chaudhuri, Mandal and Chaudhuri, (2002) have developed a page layout analyzer that can locate the textual zones in multi-column and multi-font Indian documents in Devnagari and Bengali scripts, but are not friendly with images and tables. On the other hand, in real time work (especially for the old and rare books), we noticed that the output of the OCR is also not very surprising and highly depends on the quality of the documents and is not robust in font variations. In such circumstances, the only way to give effective service to the requesting user is to send the scanned document images directly from the server via intranet and/or Internet.

The article addresses a useful technique that has much impact on electronic content management and content delivery system, an important issue of modern day digital library system of any kind. The technique presented here is easy to implement. The main idea is that, the non-text portion, say, of a book-page, such as four-side margins, space between two paragraphs and alike blank/white regions have been properly marked and discarded during the preparation phase of a derived digital document. Thus, it saves a reasonable amount of storage space in any means. A simple as well as conventional noise cleaning algorithm is used to make the ultimate image noise free. Further, when the ultimate document images are noise free it is prominent and easy to read. In essence, the technique saves a reasonable amount of storage space when in the server side of the digital library. The digital obsolesce and the management of storage space for digital library has become a burning issue as the electronic documents are going to be doubled in every six months. The detailed discussion about it is beyond the scope of this study. Obviously, the web request for such pages is served in better speed than usual process. Moreover, a margin resetting algorithm at client side has been applied to return those pages in its original form. This helps the reader to read the web document having usual margin. Sometimes, if required, the reader may set the margin according to his convenience. This option is available in the margin setting algorithm. Considering, mainly the storage space and the access speed while serving web request the obtained results have some positive implications. This indicates that this technique is much viable when a derived digital document is served through web and provides a simple probable solution towards the problem of e-content management and its delivery in digital library system. This experiment has been performed with many variants of derived digital documents and the results obtained are very much impressive. The results, point out that the conventional practice for scanning the paper document and the practice of storing the raw document pages is a pedestrian approach though the practice is still going on.

This paper is organized in following five sections. Section 2 describes the image acquisition and noise cleaning techniques. The storage optimization technique to save space in the server is presented in section 3. Some experimental results and discussion with various types of input images are shown in section 4. Finally section 5 embodies conclusion of the study.

2. Image Acquisition and Noise Cleaning

The first step of a digital library is the content collection/creation. Document image collection includes image acquisition that is, to scan a document with appropriate scanning instrument. The purpose requires an imaging sensor and the capability to digitize the signal produced by the sensor. Digitization can be done either by a flatbed scanner or a hand-held scanner. The resolution of a flatbed scanner varies from 100

to 900 dots per inch (dpi) where as the hand-held scanners typically have a lower resolution range. For the present study, the flatbed scanners are appropriate so that a complete document page can be digitized at a time, though flatbed scanners face some difficulties. Flatbed scanners often come in contact with at least that part of the object that has to be scanned. They also require books to be fully opened most of the time. Both practices can damage rare books; for example, opening a book at 180 degrees can damage its spine. A planetary scanner (also called an orbital scanner) is a type of image scanner used for making scans of rare books and other easily damaged documents. In essence, such a scanner has a mounted camera which takes photos of a well-lit environment. Originally, such scanners were expensive and could only be found in archives and museums. This system is equaly applicable for the scanned document images from the planetary scanner also. All inputs of the system are as uncompressed gray scale TIFF (Tagged Image File Format) file format and scanned with 300 dpi. Figure 1 shows the paper-based text and digitized text shown in XV image viewer software.

In the TIFF header, number of rows are available in the value field of the directory entry of Tag Value 257; number of columns are available in the value field of the directory entry of Tag Value 256, and Strip Offset from the directory entry of Tag Value 273. We use Tag Values 257 and 256 for number of rows and for number of columns but for Strip Offset = file size – (rows * columns). From Strip Offset position, every character in a 2D matrix, it is the gray-label matrix of that image.

Figure 1 (a) Paper based text (b) Digitized text shown in xv

2.1 Noise Cleaning

Binarization is a process for converting a gray-tone image to a two-tone image or binary image. In binary image, the pixels are either black (usually having value *1*) or white (usually having value *0*). Alternatively, the pixel or point (r,c) with value *1* is called object pixel and the pixel with value *0* is called background pixel.

Thresholding is the easiest method to binarize an image. Check every pixel if it is greater than or equal to a particular value T, then set that pixel with *0* else with *1*. The value T is called a threshold value. $B(x,y) = 0$ if $G(x,y) > T$ and *1* if $G(x,y) < T$. Where B(x,y) is the pixel in the binary image and G(x,y) is the pixel in the gray-label image. Threshold selection is an important task for binarization. One popular method is using histogram. In bimodal histogram, select the bottom of the valley between peaks of the histogram. The histogram with two peaks is termed as bimodal. Here, we assume that the images are bimodal, we use threshold value *128* in our binarization module. Though, there is a scope to change the threshold value. Figure 2 shows the binarized image of the image shown in Figure 1(b). In this two-tone image, the noises due to shadow is automatically removed as it has a gray value greater than the *threshold value*.

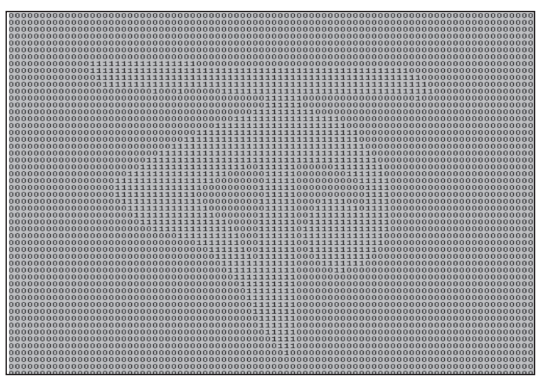
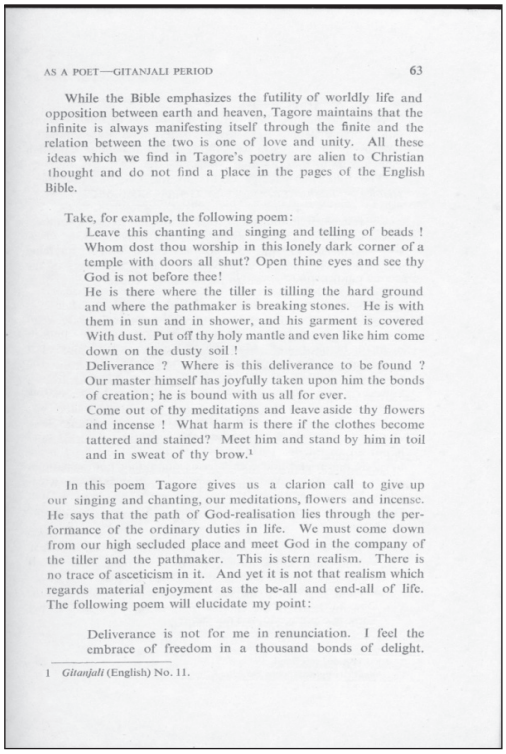


Figure 2 Binarized image

3. Storage Optimization

A close look at any book or a printed document will reveal that almost 30%- 50% of the total page size of any book are blank mainly for the margins setup on the four sides of a page of the book and more accurately some half and blank pages also increase the percentage of the blank space available normally in all books. In other words, on an average, 50%-70% part of the page area contains valid text. Our objective is to deal with the textual portion without changing the original arrangement of the page setup and store it into the server so that the text portion can be retrieved as and when required. It also reduces the document image size as it excludes the four margins. As a case study, we considered a rarely available book, '*Rabindranath Tagore His Mind and Art'* authored by B S Chakraborty, Young India Publications, New Delhi-48. The book contains 303 pages. Arbitrarily, a normal page was chosen from the book and scanned with 300dpi in a flatbed scanner. Figure 3a depicts the document image of the page with full margin whereas Figure 3b presents the document image of the same page with the text part only, without four margins. The file sizes of the two document images with (Figure 3a) and without margin (Figure 3b) are respectively 3894 KB and 2636 KB and eventually the size difference is 1258 KB. So, for every single full page one can save 1258 KB memory space. Similarly, to store 303 pages of the entire book one can save 381 MB storage space even less as normally all the books have some half and quarter pages, as described above.

But in normal convention readers are not habituated to read a book, either in the hard or soft form, without margins. Our system has been designed in such a way that the margins of the pages can be computed according to the choice



AS A POET—GITANJALI PERIOD 63

While the Bible emphasizes the futility of worldly life and opposition between earth and heaven, Tagore maintains that the infinite is always manifesting itself through the finite and the relation between the two is one of love and unity. All these ideas which we find in Tagore's poetry are alien to Christian thought and do not find a place in the pages of the English Bible.

Take, for example, the following poem:

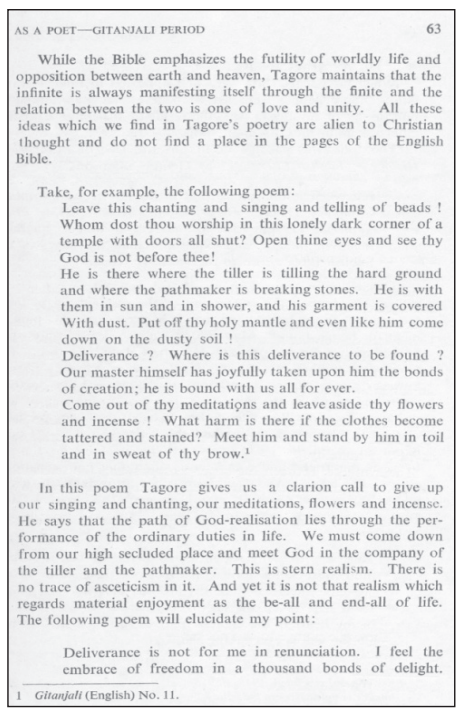
Leave this chanting and singing and telling of beads ! Whom dost thou worship in this lonely dark corner of a temple with doors all shut? Open thine eyes and see thy God is not before thee!
He is there where the tiller is tilling the hard ground and where the pathmaker is breaking stones. He is with them in sun and in shower, and his garment is covered With dust. Put off thy holy mantle and even like him come down on the dusty soil !
Deliverance ? Where is this deliverance to be found ? Our master himself has joyfully taken upon him the bonds of creation; he is bound with us all for ever.
Come out of thy meditations and leave aside thy flowers and incense ! What harm is there if the clothes become tattered and stained? Meet him and stand by him in toil and in sweat of thy brow.[1]

In this poem Tagore gives us a clarion call to give up our singing and chanting, our meditations, flowers and incense. He says that the path of God-realisation lies through the performance of the ordinary duties in life. We must come down from our high secluded place and meet God in the company of the tiller and the pathmaker. This is stern realism. There is no trace of asceticism in it. And yet it is not that realism which regards material enjoyment as the be-all and end-all of life. The following poem will elucidate my point:

Deliverance is not for me in renunciation. I feel the embrace of freedom in a thousand bonds of delight.

1 *Gitanjali* (English) No. 11.

Figure 3a The scanned page with margin



AS A POET—GITANJALI PERIOD 63

While the Bible emphasizes the futility of worldly life and opposition between earth and heaven, Tagore maintains that the infinite is always manifesting itself through the finite and the relation between the two is one of love and unity. All these ideas which we find in Tagore's poetry are alien to Christian thought and do not find a place in the pages of the English Bible.

Take, for example, the following poem:

Leave this chanting and singing and telling of beads ! Whom dost thou worship in this lonely dark corner of a temple with doors all shut? Open thine eyes and see thy God is not before thee!
He is there where the tiller is tilling the hard ground and where the pathmaker is breaking stones. He is with them in sun and in shower, and his garment is covered With dust. Put off thy holy mantle and even like him come down on the dusty soil !
Deliverance ? Where is this deliverance to be found ? Our master himself has joyfully taken upon him the bonds of creation; he is bound with us all for ever.
Come out of thy meditations and leave aside thy flowers and incense ! What harm is there if the clothes become tattered and stained? Meet him and stand by him in toil and in sweat of thy brow.[1]

In this poem Tagore gives us a clarion call to give up our singing and chanting, our meditations, flowers and incense. He says that the path of God-realisation lies through the performance of the ordinary duties in life. We must come down from our high secluded place and meet God in the company of the tiller and the pathmaker. This is stern realism. There is no trace of asceticism in it. And yet it is not that realism which regards material enjoyment as the be-all and end-all of life. The following poem will elucidate my point:

Deliverance is not for me in renunciation. I feel the embrace of freedom in a thousand bonds of delight.

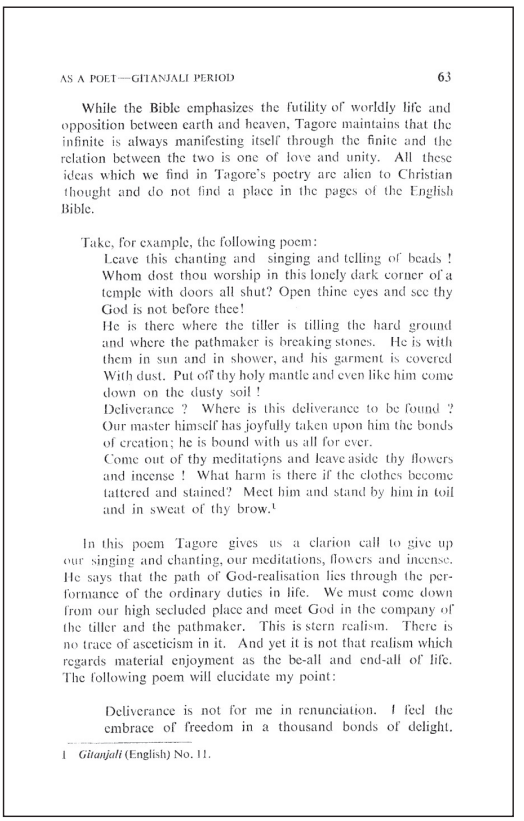
1 *Gitanjali* (English) No. 11.

Figure 3b The scanned page with only textual zone

of the reader and the reader can easily put the margin size to view the actual text according to one's need. The document is re-written in a new form after noise removal as discussed in section 3. The noise removal algorithm has been applied in the document images of Figure 3b and the outputs have been shown in Figures 4a and 4b with margins 1 inch and 1.5 inches respectively. The retrieval algorithm also works a bit faster than usual as its performance depends mainly on the file size. As a result, the electronic document serves the requests through the network in a better speed and in a more readable form. The program for re-writing the images with the desired margin may reside at server-side or in the client-side. Though, for both the cases there are some advantages and disadvantages and our algorithm is equally applicable for both the cases and the choice between the two is a web-designing issue only.

4. Experimental Results

An efficient as well as easy-to-implement technique has been proposed in this study for the management and delivery of e-content in the context of digital library. The proposed method has been applied with many variants of document images having different languages, font size and documents having possibility of containing considerable amount of noise due to old age and careless holding of the original documents. Sample of original scanned images and the obtained results have been presented in the Figures 5a, 5b, Figures 6a, 6b and Figures 7a, 7b respectively and the detailed analysis of



AS A POET—GITANJALI PERIOD 63

While the Bible emphasizes the futility of worldly life and opposition between earth and heaven, Tagore maintains that the infinite is always manifesting itself through the finite and the relation between the two is one of love and unity. All these ideas which we find in Tagore's poetry are alien to Christian thought and do not find a place in the pages of the English Bible.

Take, for example, the following poem:

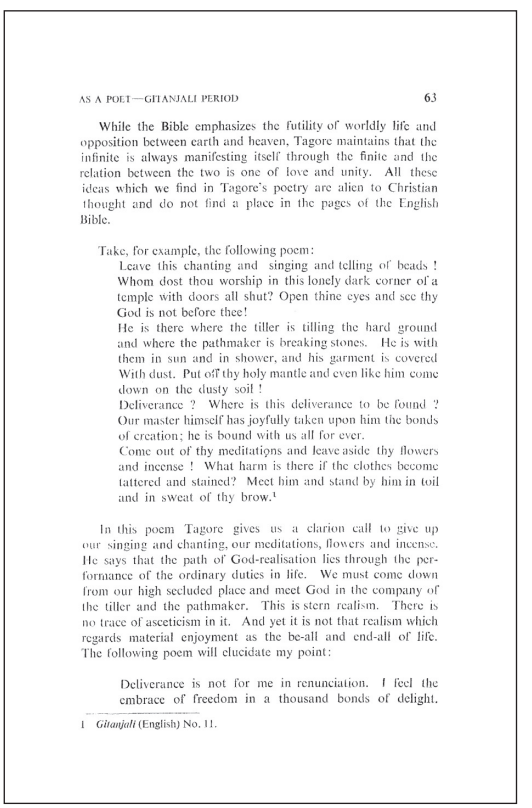
Leave this chanting and singing and telling of beads ! Whom dost thou worship in this lonely dark corner of a temple with doors all shut? Open thine eyes and see thy God is not before thee!
He is there where the tiller is tilling the hard ground and where the pathmaker is breaking stones. He is with them in sun and in shower, and his garment is covered With dust. Put off thy holy mantle and even like him come down on the dusty soil !
Deliverance ? Where is this deliverance to be found ? Our master himself has joyfully taken upon him the bonds of creation; he is bound with us all for ever.
Come out of thy meditations and leave aside thy flowers and incense ! What harm is there if the clothes become tattered and stained? Meet him and stand by him in toil and in sweat of thy brow.[1]

In this poem Tagore gives us a clarion call to give up our singing and chanting, our meditations, flowers and incense. He says that the path of God-realisation lies through the performance of the ordinary duties in life. We must come down from our high secluded place and meet God in the company of the tiller and the pathmaker. This is stern realism. There is no trace of asceticism in it. And yet it is not that realism which regards material enjoyment as the be-all and end-all of life. The following poem will elucidate my point:

Deliverance is not for me in renunciation. I feel the embrace of freedom in a thousand bonds of delight.

1 *Gitanjali* (English) No. 11.

Figure 4a The output image with 1″ margin.



AS A POET—GITANJALI PERIOD 63

While the Bible emphasizes the futility of worldly life and opposition between earth and heaven, Tagore maintains that the infinite is always manifesting itself through the finite and the relation between the two is one of love and unity. All these ideas which we find in Tagore's poetry are alien to Christian thought and do not find a place in the pages of the English Bible.

Take, for example, the following poem:

Leave this chanting and singing and telling of beads ! Whom dost thou worship in this lonely dark corner of a temple with doors all shut? Open thine eyes and see thy God is not before thee!
He is there where the tiller is tilling the hard ground and where the pathmaker is breaking stones. He is with them in sun and in shower, and his garment is covered With dust. Put off thy holy mantle and even like him come down on the dusty soil !
Deliverance ? Where is this deliverance to be found ? Our master himself has joyfully taken upon him the bonds of creation; he is bound with us all for ever.
Come out of thy meditations and leave aside thy flowers and incense ! What harm is there if the clothes become tattered and stained? Meet him and stand by him in toil and in sweat of thy brow.[1]

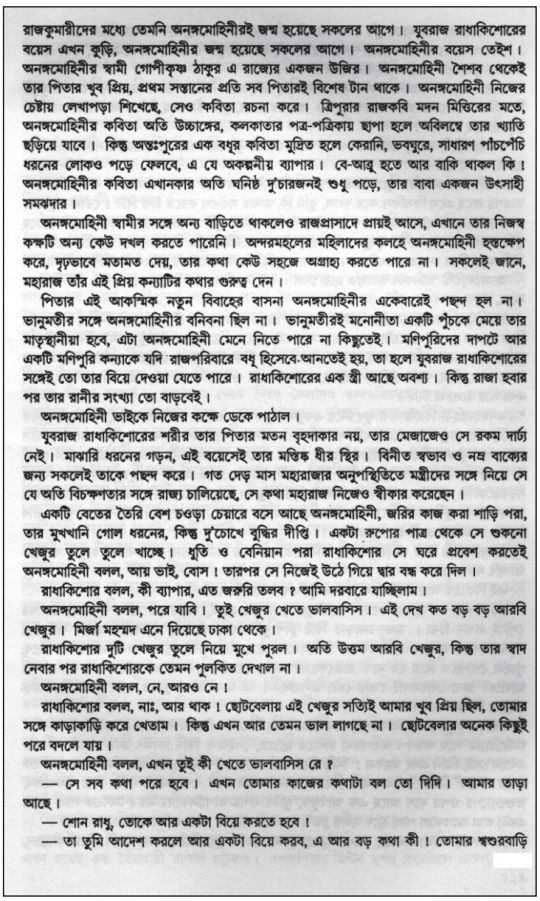
In this poem Tagore gives us a clarion call to give up our singing and chanting, our meditations, flowers and incense. He says that the path of God-realisation lies through the performance of the ordinary duties in life. We must come down from our high secluded place and meet God in the company of the tiller and the pathmaker. This is stern realism. There is no trace of asceticism in it. And yet it is not that realism which regards material enjoyment as the be-all and end-all of life. The following poem will elucidate my point:

Deliverance is not for me in renunciation. I feel the embrace of freedom in a thousand bonds of delight.

1 *Gitanjali* (English) No. 11.

Figure 4b The output image with 1.5″ margin.

the results are presented in Table 1. Figure 5a is the scanned image of the original documents written in Bengali script and Figure 5b is the noise-free output image obtained with the application of the proposed method. The four side margins of the page have been set to 1 inch, here. Qualitatively, with the normal vision it is seen that the original scanned image (Figure 5a) has some obvious noise, which may have occurred due to back page shadow or something else. From Table 1 it is seen that, the input image shown in Fig. 5a has the file size of 3754 KB where as processed image (Fig. 5.b) is of 5040 KB. The four-sided white margins have been set with 1-inch width. The percentage of reduction of space here is 34. It should be noted that, we are accessing a page of size 5040 KB for which the original file size is 3754 KB stored in the server taking only the textual regions. Similarly Fig. 6.a and Fig. 6.b are examples of another document image of different type but printed in Bengali. In this case the margin set up is of 1.5 inch and the percentage of space saved is 61. Fig. 7a is a sample document written in Devnagari script. Here, the reduction of space has also occurred. It is important to note that the proposed method is script independent. Considering the amount of the total size of a document image of all pages of a book, it can be concluded that the saved space



রাজকুমারীদের মধ্যে তেমনি অনঙ্গমোহিনীরই জন্ম হয়েছে সকলের আগে। যুবরাজ রাধাকিশোরের বয়েস এখন কুড়ি, অনঙ্গমোহিনীর জন্ম হয়েছে সকলের আগে। অনঙ্গমোহিনীর বয়েস তেইশ। অনঙ্গমোহিনীর স্বামী গোপীকৃষ্ণ ঠাকুর এ রাজ্যের একজন উজির। অনঙ্গমোহিনী শৈশব থেকেই তার পিতার খুব প্রিয়, প্রথম সন্তানের প্রতি সব পিতারই বিশেষ টান থাকে। অনঙ্গমোহিনী নিজের চেষ্টায় লেখাপড়া শিখেছে, সেও কবিতা রচনা করে। ত্রিপুরার রাজকবি মদন মিত্তিরের মতে, অনঙ্গমোহিনীর কবিতা অতি উচ্চাঙ্গের, কলকাতার পত্র-পত্রিকায় ছাপা হলে অবিলম্বে তার খ্যাতি ছড়িয়ে যাবে। কিন্তু অন্তঃপুরের এক বধূর কবিতা মুদ্রিত হলে কেরানি, ভবঘুরে, সাধারণ পাঁচপেঁচি ধরনের লোকও পড়ে ফেলবে, এ যে অকল্পনীয় ব্যাপার। কে-আর্ হতে আর বাকি থাকল কি। অনঙ্গমোহিনীর কবিতা এখানকার অতি ঘনিষ্ঠ দু'চারজনই শুধু পড়ে, তার বাবা একজন উৎসাহী সমঝদার।

অনঙ্গমোহিনী স্বামীর সঙ্গে অন্য বাড়িতে থাকলেও রাজপ্রাসাদে প্রায়ই আসে, এখানে তার নিজস্ব কক্ষটি অন্য কেউ দখল করতে পারেনি। অন্দরমহলের মহিলাদের কলহে অনঙ্গমোহিনী হস্তক্ষেপ করে, দৃঢ়ভাবে মতামত দেয়, তার কথা কেউ সহজে অগ্রাহ্য করতে পারে না। সকলেই জানে, মহারাজ তাঁর এই প্রিয় কন্যাটির কথার গুরুত্ব দেন।

পিতার এই আকস্মিক নতুন বিবাহের বাসনা অনঙ্গমোহিনীর একেবারেই পছন্দ হল না। ভানুমতীর সঙ্গে অনঙ্গমোহিনীর বনিবনা ছিল না। ভানুমতীরই মনোনীতা একটি পুঁচকে মেয়ে তার মাতৃস্থানীয়া হবে, এটা অনঙ্গমোহিনী মেনে নিতে পারে না কিছুতেই। মণিপুরিদের দাপটে আর একটি মণিপুরি কন্যাকে যদি রাজপরিবারে বধূ হিসেবে আনতেই হয়, তা হলে যুবরাজ রাধাকিশোরের সঙ্গেই তো তার বিয়ে দেওয়া যেতে পারে। রাধাকিশোরের এক স্ত্রী আছে অবশ্য। কিন্তু রাজা হবার পর তার রানীর সংখ্যা তো বাড়বেই।

অনঙ্গমোহিনী ভাইকে নিজের কক্ষে ডেকে পাঠাল।

যুবরাজ রাধাকিশোরের শরীর তার পিতার মতন বৃহদাকার নয়, তার মেজাজেও সে রকম দার্ঢ্য নেই। মাঝারি ধরনের গড়ন, এই বয়েসেই তার মস্তিষ্ক ধীর স্থির। বিনীত স্বভাব ও নম্র বাক্যের জন্য সকলেই তাকে পছন্দ করে। গত দেড় মাস মহারাজার অনুপস্থিতিতে মন্ত্রীদের সঙ্গে নিয়ে সে যে অতি বিচক্ষণতার সঙ্গে রাজ্য চালিয়েছে, সে কথা মহারাজ নিজেও স্বীকার করেছেন।

একটি বেতের তৈরি বেশ চওড়া চেয়ারে বসে আছে অনঙ্গমোহিনী, জরির কাজ করা শাড়ি পরা, তার মুখখানি গোল ধরনের, কিন্তু দু'চোখে বুদ্ধির দীপ্তি। একটা রুপোর পাত্র থেকে সে শুকনো খেজুর তুলে তুলে খাচ্ছে। ধুতি ও বেনিয়ান পরা রাধাকিশোর সে ঘরে প্রবেশ করতেই অনঙ্গমোহিনী বলল, আয় ভাই, বোস। তারপর সে নিজেই উঠে গিয়ে দ্বার বন্ধ করে দিল।

রাধাকিশোর বলল, কী ব্যাপার, এত জরুরি তলব ? আমি দরবারে যাচ্ছিলাম।

অনঙ্গমোহিনী বলল, পরে যাবি। তুই খেজুর খেতে ভালবাসিস। এই দেখ কত বড় বড় আরবি খেজুর। মির্জা মহম্মদ এনে দিয়েছে ঢাকা থেকে।

রাধাকিশোর দুটি খেজুর তুলে নিয়ে মুখে পুরল। অতি উত্তম আরবি খেজুর, কিন্তু তার স্বাদ নেবার পর রাধাকিশোরকে তেমন পুলকিত দেখাল না।

অনঙ্গমোহিনী বলল, নে, আরও নে।

রাধাকিশোর বলল, নাঃ, আর থাক। ছোটবেলায় এই খেজুর সত্যিই আমার খুব প্রিয় ছিল, তোমার সঙ্গে কাড়াকাড়ি করে খেতাম। কিন্তু এখন আর তেমন ভাল লাগছে না। ছোটবেলার অনেক কিছুই পরে বদলে যায়।

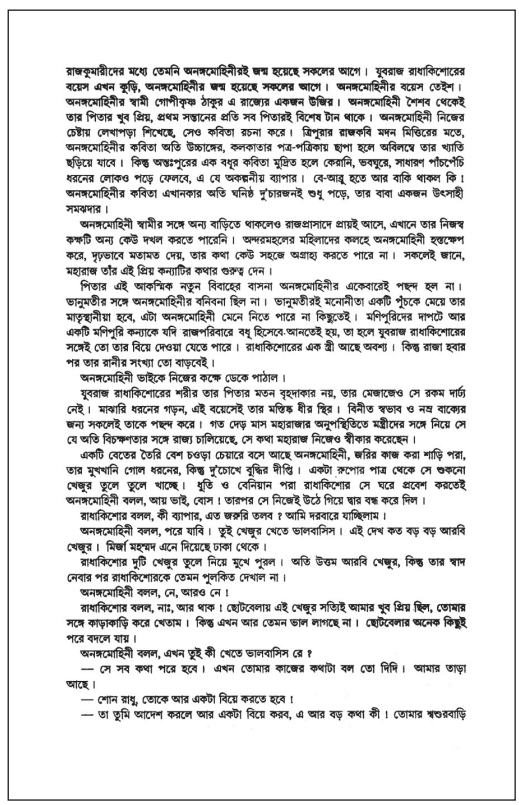
অনঙ্গমোহিনী বলল, এখন তুই কী খেতে ভালবাসিস রে ?

— সে সব কথা পরে হবে। এখন তোমার কাজের কথাটা বল তো দিদি। আমার তাড়া আছে।

— শোন রাধু, তোকে আর একটা বিয়ে করতে হবে !

— তা তুমি আদেশ করলে আর একটা বিয়ে করব, এ আর বড় কথা কী ! তোমার শ্বশুরবাড়ি

Figure 5a The original scanned page.



রাজকুমারীদের মধ্যে তেমনি অনঙ্গমোহিনীরই জন্ম হয়েছে সকলের আগে। যুবরাজ রাধাকিশোরের বয়েস এখন কুড়ি, অনঙ্গমোহিনীর জন্ম হয়েছে সকলের আগে। অনঙ্গমোহিনীর বয়েস তেইশ। অনঙ্গমোহিনীর স্বামী গোপীকৃষ্ণ ঠাকুর এ রাজ্যের একজন উজির। অনঙ্গমোহিনী শৈশব থেকেই তার পিতার খুব প্রিয়, প্রথম সন্তানের প্রতি সব পিতারই বিশেষ টান থাকে। অনঙ্গমোহিনী নিজের চেষ্টায় লেখাপড়া শিখেছে, সেও কবিতা রচনা করে। ত্রিপুরার রাজকবি মদন মিত্তিরের মতে, অনঙ্গমোহিনীর কবিতা অতি উচ্চাঙ্গের, কলকাতার পত্র-পত্রিকায় ছাপা হলে অবিলম্বে তার খ্যাতি ছড়িয়ে যাবে। কিন্তু অন্তঃপুরের এক বধূর কবিতা মুদ্রিত হলে কেরানি, ভবঘুরে, সাধারণ পাঁচপেঁচি ধরনের লোকও পড়ে ফেলবে, এ যে অকল্পনীয় ব্যাপার। কে-আর্ হতে আর বাকি থাকল কি। অনঙ্গমোহিনীর কবিতা এখানকার অতি ঘনিষ্ঠ দু'চারজনই শুধু পড়ে, তার বাবা একজন উৎসাহী সমঝদার।

অনঙ্গমোহিনী স্বামীর সঙ্গে অন্য বাড়িতে থাকলেও রাজপ্রাসাদে প্রায়ই আসে, এখানে তার নিজস্ব কক্ষটি অন্য কেউ দখল করতে পারেনি। অন্দরমহলের মহিলাদের কলহে অনঙ্গমোহিনী হস্তক্ষেপ করে, দৃঢ়ভাবে মতামত দেয়, তার কথা কেউ সহজে অগ্রাহ্য করতে পারে না। সকলেই জানে, মহারাজ তাঁর এই প্রিয় কন্যাটির কথার গুরুত্ব দেন।

পিতার এই আকস্মিক নতুন বিবাহের বাসনা অনঙ্গমোহিনীর একেবারেই পছন্দ হল না। ভানুমতীর সঙ্গে অনঙ্গমোহিনীর বনিবনা ছিল না। ভানুমতীরই মনোনীতা একটি পুঁচকে মেয়ে তার মাতৃস্থানীয়া হবে, এটা অনঙ্গমোহিনী মেনে নিতে পারে না কিছুতেই। মণিপুরিদের দাপটে আর একটি মণিপুরি কন্যাকে যদি রাজপরিবারে বধূ হিসেবে আনতেই হয়, তা হলে যুবরাজ রাধাকিশোরের সঙ্গেই তো তার বিয়ে দেওয়া যেতে পারে। রাধাকিশোরের এক স্ত্রী আছে অবশ্য। কিন্তু রাজা হবার পর তার রানীর সংখ্যা তো বাড়বেই।

অনঙ্গমোহিনী ভাইকে নিজের কক্ষে ডেকে পাঠাল।

যুবরাজ রাধাকিশোরের শরীর তার পিতার মতন বৃহদাকার নয়, তার মেজাজেও সে রকম দার্ঢ্য নেই। মাঝারি ধরনের গড়ন, এই বয়েসেই তার মস্তিষ্ক ধীর স্থির। বিনীত স্বভাব ও নম্র বাক্যের জন্য সকলেই তাকে পছন্দ করে। গত দেড় মাস মহারাজার অনুপস্থিতিতে মন্ত্রীদের সঙ্গে নিয়ে সে যে অতি বিচক্ষণতার সঙ্গে রাজ্য চালিয়েছে, সে কথা মহারাজ নিজেও স্বীকার করেছেন।

একটি বেতের তৈরি বেশ চওড়া চেয়ারে বসে আছে অনঙ্গমোহিনী, জরির কাজ করা শাড়ি পরা, তার মুখখানি গোল ধরনের, কিন্তু দু'চোখে বুদ্ধির দীপ্তি। একটা রুপোর পাত্র থেকে সে শুকনো খেজুর তুলে তুলে খাচ্ছে। ধুতি ও বেনিয়ান পরা রাধাকিশোর সে ঘরে প্রবেশ করতেই অনঙ্গমোহিনী বলল, আয় ভাই, বোস। তারপর সে নিজেই উঠে গিয়ে দ্বার বন্ধ করে দিল।

রাধাকিশোর বলল, কী ব্যাপার, এত জরুরি তলব ? আমি দরবারে যাচ্ছিলাম।

অনঙ্গমোহিনী বলল, পরে যাবি। তুই খেজুর খেতে ভালবাসিস। এই দেখ কত বড় বড় আরবি খেজুর। মির্জা মহম্মদ এনে দিয়েছে ঢাকা থেকে।

রাধাকিশোর দুটি খেজুর তুলে নিয়ে মুখে পুরল। অতি উত্তম আরবি খেজুর, কিন্তু তার স্বাদ নেবার পর রাধাকিশোরকে তেমন পুলকিত দেখাল না।

অনঙ্গমোহিনী বলল, নে, আরও নে।

রাধাকিশোর বলল, নাঃ, আর থাক। ছোটবেলায় এই খেজুর সত্যিই আমার খুব প্রিয় ছিল, তোমার সঙ্গে কাড়াকাড়ি করে খেতাম। কিন্তু এখন আর তেমন ভাল লাগছে না। ছোটবেলার অনেক কিছুই পরে বদলে যায়।

অনঙ্গমোহিনী বলল, এখন তুই কী খেতে ভালবাসিস রে ?

— সে সব কথা পরে হবে। এখন তোমার কাজের কথাটা বল তো দিদি। আমার তাড়া আছে।

— শোন রাধু, তোকে আর একটা বিয়ে করতে হবে।

— তা তুমি আদেশ করলে আর একটা বিয়ে করব, এ আর বড় কথা কী। তোমার শ্বশুরবাড়ি

Figure 5b The output image with 1″ margin.

বৈদ্যুতিন অভিধানের সম্ভাবনা

বিদ্যুৎবরণ চৌধুরী

মানুষের শিক্ষালাভ ও জ্ঞানবৃদ্ধির উপায়গুলি মোটামুটি দুভাগে ভাগ করা যায়। তার একটি পাঠ্য ও শ্রাব্য ভাষানির্ভর, অন্যটি ভাষাবহির্ভূত। পাঠ্য ভাষাশ্রিত বিষয়গুলি বর্ণমালায় লিপিবদ্ধ বা মুদ্রিত হয়। যেমন বইপত্র, নোটিস, রিপোর্ট, জার্নাল, মিসিল, পত্রপত্রিকা, খবরের কাগজ। অন্যদিকে ক্লাস, সেমিনার, আলোচনা, বিতর্ক ও নানা ধরনের বক্তৃতার মাধ্যমে আমরা শ্রাব্য উপায়ে জ্ঞান অর্জন করি। ভাষাবহির্ভূত উপায়গুলি প্রধানত দৃষ্টিনির্ভর বা দৃশ্য। অবশ্য স্পর্শ, ঘ্রাণ জাতীয় কিছু ইন্দ্রিয়গ্রাহ্য যৌগ উপায়ও আছে।

আধুনিক বিজ্ঞান এবং প্রযুক্তি পাঠ্য ও শ্রাব্য এই দুই মাধ্যমেই নতুন মাত্রা যোগ করেছে। পাঠ্যবস্তু প্রকাশনা দ্রুত ও সুলভ হয়েছে, সংবাদ এবং তথ্যের আদানপ্রদান হয়েছে দ্রুত ও নির্ভরযোগ্য। রেডিও, টেলিভিশন, টেপরেকর্ডার ইত্যাদি যেসব নতুন প্রযুক্তি বাজারে এসেছে, তাদের চিত্তবিনোদন ছাড়াও শিক্ষামূলক একটা ভূমিকা আছে। যেমন টেপরেকর্ডার। দৃষ্টিহীনদের জন্য লাইব্রেরিতে রেইল পদ্ধতিতে প্রস্তুত বইয়ের স্থান এখন অনেকাংশে দখল করেছে অডিও ক্যাসেট। ক্যাসেটে টেপ করে রাখা পাঠ্যবস্তু টেপরেকর্ডারে বাজিয়ে দৃষ্টিহীন শিক্ষার্থীরা পড়াশোনা করতে পারেন।

সম্প্রতি একটি অতি শক্তিশালী মাধ্যম মানবসভ্যতা ও সংস্কৃতির প্রায় সর্বক্ষেত্রে প্রভাব বিস্তার করেছে। বছর খানেক আগে কম্পিউটারের সঙ্গে জুড় এই মাধ্যমের নাম ইলেকট্রনিক মিডিয়া। কম্পিউটারের প্রভাব বিশ্ব শতাব্দীর দ্বিতীয়ার্ধ থেকে অনুভব করা গেলেও তার জনপ্রিয়তা শুরু হয়েছে ব্যক্তিগত কম্পিউটার বা পি-সির উদ্ভাবনের পর থেকে। শুরুতে পি-সি ছিল যাচ্ছেই দামি, সুতরাং ভারত্রীয় মধ্যবিত্তের নাগালের বাইরে। তাছাড়া এর জন্য শীততাপনিয়ন্ত্রিত ঘর, ভোল্টেজ স্টেবিলাইজার ইত্যাদির প্রয়োজন হত। বছর দশেক হল পি-সি শস্তা, শক্তিশালী ও সহজে ব্যবহারযোগ্য হয়েছে।

শুধু শক্তিশালী হওয়াই নয়, পি-সি এখন আকারেও ক্রমশ ছোট হচ্ছে। একটা অ্যাটাচিতে পুরে যেখানে সেখানে নিয়ে যাওয়ার জন্য বেরিয়েছে ল্যাপ-টপ পি-সি। কোলের ওপর রেখে ব্যবহার করা যায় বলে এরকম নাম। কিছুদিন হল ল্যাপ-টপের চেয়েও ছোট পি-সি বাজারে দেখা যাচ্ছে। পাম-টপ। এক হাতের তালুর ওপর রেখে অন্য হাতে চালানো যায়, এত ছোট।

কম্পিউটার, বিশেষ করে পি-সি আসার পর থেকে পাঠ্যবস্তুর দুধরনের বহিরাঙ্গিক রূপ দেখা যাচ্ছে। একটি রূপ অ্যাসকি চিহ্নপরিচিত কাগজে লেখা বা ছাপা বইপত্র, ইলেকট্রনিক পরিভাষায় যাকে বলা যায় 'সফ্ট প্রতিলিপি' বা হার্ড কপি। অন্যদিকে আছে 'নন্ প্রতিলিপি' বা সফ্ট কপি, যা পি-সির অভ্যন্তরে হার্ড-ডিস্ক মেমরিতে রেখে দেওয়া যায়

Figure 6a The original scanned page with font variation.

বৈদ্যুতিন অভিধানের সম্ভাবনা

বিদ্যুৎবরণ চৌধুরী

মানুষের শিক্ষালাভ ও জ্ঞানবৃদ্ধির উপায়গুলি মোটামুটি দুভাগে ভাগ করা যায়। তার একটি পাঠ্য ও শ্রাব্য ভাষানির্ভর, অন্যটি ভাষাবহির্ভূত। পাঠ্য ভাষাশ্রিত বিষয়গুলি বর্ণমালায় লিপিবদ্ধ বা মুদ্রিত হয়। যেমন বইপত্র, নোটিস, রিপোর্ট, জার্নাল, মিসিল, পত্রপত্রিকা, খবরের কাগজ। অন্যদিকে ক্লাস, সেমিনার, আলোচনা, বিতর্ক ও নানা ধরনের বক্তৃতার মাধ্যমে আমরা শ্রাব্য উপায়ে জ্ঞান অর্জন করি। ভাষাবহির্ভূত উপায়গুলি প্রধানত দৃষ্টিনির্ভর বা দৃশ্য। অবশ্য স্পর্শ, ঘ্রাণ জাতীয় কিছু ইন্দ্রিয়গ্রাহ্য যৌগ উপায়ও আছে।

আধুনিক বিজ্ঞান এবং প্রযুক্তি পাঠ্য ও শ্রাব্য এই দুই মাধ্যমেই নতুন মাত্রা যোগ করেছে। পাঠ্যবস্তু প্রকাশনা দ্রুত ও সুলভ হয়েছে, সংবাদ এবং তথ্যের আদানপ্রদান হয়েছে দ্রুত ও নির্ভরযোগ্য। রেডিও, টেলিভিশন, টেপরেকর্ডার ইত্যাদি যেসব নতুন প্রযুক্তি বাজারে এসেছে, তাদের চিত্তবিনোদন ছাড়াও শিক্ষামূলক একটা ভূমিকা আছে। যেমন টেপরেকর্ডার। দৃষ্টিহীনদের জন্য লাইব্রেরিতে রেইল পদ্ধতিতে প্রস্তুত বইয়ের স্থান এখন অনেকাংশে দখল করেছে অডিও ক্যাসেট। ক্যাসেটে টেপ করে রাখা পাঠ্যবস্তু টেপরেকর্ডারে বাজিয়ে দৃষ্টিহীন শিক্ষার্থীরা পড়াশোনা করতে পারেন।

সম্প্রতি একটি অতি শক্তিশালী মাধ্যম মানবসভ্যতা ও সংস্কৃতির প্রায় সর্বক্ষেত্রে প্রভাব বিস্তার করেছে। বছর খানেক আগে কম্পিউটারের সঙ্গে জুড় এই মাধ্যমের নাম ইলেকট্রনিক মিডিয়া। কম্পিউটারের প্রভাব বিশ্ব শতাব্দীর দ্বিতীয়ার্ধ থেকে অনুভব করা গেলেও তার জনপ্রিয়তা শুরু হয়েছে ব্যক্তিগত কম্পিউটার বা পি-সির উদ্ভাবনের পর থেকে। শুরুতে পি-সি ছিল যাচ্ছেই দামি, সুতরাং ভারত্রীয় মধ্যবিত্তের নাগালের বাইরে। তাছাড়া এর জন্য শীততাপনিয়ন্ত্রিত ঘর, ভোল্টেজ স্টেবিলাইজার ইত্যাদির প্রয়োজন হত। বছর দশেক হল পি-সি শস্তা, শক্তিশালী ও সহজে ব্যবহারযোগ্য হয়েছে।

শুধু শক্তিশালী হওয়াই নয়, পি-সি এখন আকারেও ক্রমশ ছোট হচ্ছে। একটা অ্যাটাচিতে পুরে যেখানে সেখানে নিয়ে যাওয়ার জন্য বেরিয়েছে ল্যাপ-টপ পি-সি। কোলের ওপর রেখে ব্যবহার করা যায় বলে এরকম নাম। কিছুদিন হল ল্যাপ-টপের চেয়েও ছোট পি-সি বাজারে দেখা যাচ্ছে। পাম-টপ। এক হাতের তালুর ওপর রেখে অন্য হাতে চালানো যায়, এত ছোট।

কম্পিউটার, বিশেষ করে পি-সি আসার পর থেকে পাঠ্যবস্তুর দুধরনের বহিরাঙ্গিক রূপ দেখা যাচ্ছে। একটি রূপ অ্যাসকি চিহ্নপরিচিত কাগজে লেখা বা ছাপা বইপত্র, ইলেকট্রনিক পরিভাষায় যাকে বলা যায় 'সফ্ট প্রতিলিপি' বা হার্ড কপি। অন্যদিকে আছে 'নন্ প্রতিলিপি' বা সফ্ট কপি, যা পি-সির অভ্যন্তরে হার্ড-ডিস্ক মেমরিতে রেখে দেওয়া যায়

Figure 6b The output image with 1.5″ margin.

पूंजीवादी देशों में किसानों और दस्तकारों के विपुल बहुलांश की तबाही और कंगाली को भी ध्यान में रखा जाना चाहिए।

पूंजीवादी देशों में मेहनतकशों की अवस्था में निरपेक्ष गिरावट के पीछे जो कारक हैं, उनमें से कुछ संक्षेप में ये हैं।

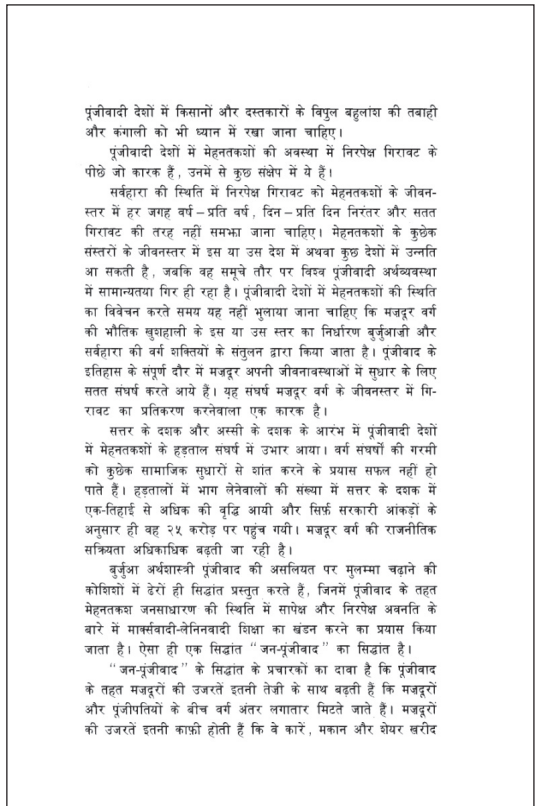
सर्वहारा की स्थिति में निरपेक्ष गिरावट को मेहनतकशों के जीवन-स्तर में हर जगह वर्ष – प्रति वर्ष, दिन – प्रति दिन निरंतर और सतत गिरावट की तरह नहीं समझा जाना चाहिए। मेहनतकशों के कुछेक संस्तरों के जीवनस्तर में इस या उस देश में अथवा कुछ देशों में उन्नति आ सकती है, जबकि वह समूचे तौर पर विश्व पूंजीवादी अर्थव्यवस्था में सामान्यतया गिर ही रहा है। पूंजीवादी देशों में मेहनतकशों की स्थिति का विवेचन करते समय यह नहीं भुलाया जाना चाहिए कि मजदूर वर्ग की भौतिक खुशहाली के इस या उस स्तर का निर्धारण बुर्जुआजी और सर्वहारा की वर्ग शक्तियों के संतुलन द्वारा किया जाता है। पूंजीवाद के इतिहास के संपूर्ण दौर में मजदूर अपनी जीवनावस्थाओं में सुधार के लिए सतत संघर्ष करते आये हैं। यह संघर्ष मजदूर वर्ग के जीवनस्तर में गिरावट का प्रतिकरण करनेवाला एक कारक है।

सत्तर के दशक और अस्सी के दशक के आरंभ में पूंजीवादी देशों में मेहनतकशों के हड़ताल संघर्ष में उभार आया। वर्ग संघर्षों की गरमी को कुछेक सामाजिक सुधारों से शांत करने के प्रयास सफल नहीं हो पाते हैं। हड़तालों में भाग लेनेवालों की संख्या में सत्तर के दशक में एक-तिहाई से अधिक की वृद्धि आयी और सिर्फ़ सरकारी आंकड़ों के अनुसार ही वह २५ करोड़ पर पहुंच गयी। मजदूर वर्ग की राजनीतिक सक्रियता अधिकाधिक बढ़ती जा रही है।

बुर्जुआ अर्थशास्त्री पूंजीवाद की असलियत पर मुलम्मा चढ़ाने की कोशिशों में ढेरों ही सिद्धांत प्रस्तुत करते हैं, जिनमें पूंजीवाद के तहत मेहनतकश जनसाधारण की स्थिति में सापेक्ष और निरपेक्ष अवनति के बारे में मार्क्सवादी-लेनिनवादी शिक्षा का खंडन करने का प्रयास किया जाता है। ऐसा ही एक सिद्धांत "जन-पूंजीवाद" का सिद्धांत है।

"जन-पूंजीवाद" के सिद्धांत के प्रचारकों का दावा है कि पूंजीवाद के तहत मजदूरों की उजरतें इतनी तेजी के साथ बढ़ती हैं कि मजदूरों और पूंजीपतियों के बीच वर्ग अंतर लगातार मिटते जाते हैं। मजदूरों की उजरतें इतनी काफ़ी होती हैं कि वे कारें, मकान और शेयर खरीद

Figure 7a The original scanned page from Hindi Page.



पूंजीवादी देशों में किसानों और दस्तकारों के विपुल बहुलांश की तबाही और कंगाली को भी ध्यान में रखा जाना चाहिए।

पूंजीवादी देशों में मेहनतकशों की अवस्था में निरपेक्ष गिरावट के पीछे जो कारक हैं, उनमें से कुछ संक्षेप में ये हैं।

सर्वहारा की स्थिति में निरपेक्ष गिरावट को मेहनतकशों के जीवन-स्तर में हर जगह वर्ष – प्रति वर्ष, दिन – प्रति दिन निरंतर और सतत गिरावट की तरह नहीं समझा जाना चाहिए। मेहनतकशों के कुछेक संस्तरों के जीवनस्तर में इस या उस देश में अथवा कुछ देशों में उन्नति आ सकती है, जबकि वह समूचे तौर पर विश्व पूंजीवादी अर्थव्यवस्था में सामान्यतया गिर ही रहा है। पूंजीवादी देशों में मेहनतकशों की स्थिति का विवेचन करते समय यह नहीं भुलाया जाना चाहिए कि मजदूर वर्ग की भौतिक खुशहाली के इस या उस स्तर का निर्धारण बुर्जुआजी और सर्वहारा की वर्ग शक्तियों के संतुलन द्वारा किया जाता है। पूंजीवाद के इतिहास के संपूर्ण दौर में मजदूर अपनी जीवनावस्थाओं में सुधार के लिए सतत संघर्ष करते आये हैं। यह संघर्ष मजदूर वर्ग के जीवनस्तर में गिरावट का प्रतिकरण करनेवाला एक कारक है।

सत्तर के दशक और अस्सी के दशक के आरंभ में पूंजीवादी देशों में मेहनतकशों के हड़ताल संघर्ष में उभार आया। वर्ग संघर्षों की गरमी को कुछेक सामाजिक सुधारों से शांत करने के प्रयास सफल नहीं हो पाते हैं। हड़तालों में भाग लेनेवालों की संख्या में सत्तर के दशक में एक-तिहाई से अधिक की वृद्धि आयी और सिर्फ़ सरकारी आंकड़ों के अनुसार ही वह २५ करोड़ पर पहुंच गयी। मजदूर वर्ग की राजनीतिक सक्रियता अधिकाधिक बढ़ती जा रही है।

बुर्जुआ अर्थशास्त्री पूंजीवाद की असलियत पर मुलम्मा चढ़ाने की कोशिशों में ढेरों ही सिद्धांत प्रस्तुत करते हैं, जिनमें पूंजीवाद के तहत मेहनतकश जनसाधारण की स्थिति में सापेक्ष और निरपेक्ष अवनति के बारे में मार्क्सवादी-लेनिनवादी शिक्षा का खंडन करने का प्रयास किया जाता है। ऐसा ही एक सिद्धांत "जन-पूंजीवाद" का सिद्धांत है।

"जन-पूंजीवाद" के सिद्धांत के प्रचारकों का दावा है कि पूंजीवाद के तहत मजदूरों की उजरतें इतनी तेजी के साथ बढ़ती हैं कि मजदूरों और पूंजीपतियों के बीच वर्ग अंतर लगातार मिटते जाते हैं। मजदूरों की उजरतें इतनी काफ़ी होती हैं कि वे कारें, मकान और शेयर खरीद

Figure 7b The output image with 1.5″ margin

Table 1 Analysis of the obtained results from the proposed technique with few sample document images.

Images	File Size (KB) (Textual zone) [Figure a]	File Size (KB) (Textual zone) [Figure b]	Save in file size (KB)	Margin size (inch)	% of the save in file size
Figure 5	3754	5040	1286	1	34
Figure 6	2889	4655	1766	1.5	61
Figure 7	2627	4335	1708	1.5	65

is of considerable amount, more importantly the output image is noise free and can be margined accordingly as well at the user side.

5. Conclusions

Digital libraries are emerging technologies for document management. It has many branches to be taken care of with equal respect for the overall development of the digital library. Proper and secured content management and delivery is one of the key issues of the digital library. More and more digital material is added every day to the digital library server. Content management technologies will be a big thing in the future. Moreover, the everyday problems of digital library help in the long term preservation of digital objects; copyright of digital material; good solutions for micro charging, and pay per view etc. All these are related more or less with the proper storage of content and planning of storage device management. Due to these problems the present status of the digital library in the third world countries is in a nascent stage.

In the present study, a simple as well as easy-to-implement approach has been presented for the storage management and delivery of e-content materials related to digital library. One of the attractive features of this approach is that it can reduce the space of any scanned document image to a considerable amount and can also in principle be applied to any digital document. Further more the noise reduction algorithm becomes very effective for this type of electronic document delivery. One may say that there are many other noise cleaning algorithms available in literature, 'Gonzalez *and* Woods' *(2004), Weeks (2007),* but for the present purpose the effectiveness of our noise-cleaning algorithm is sufficient enough. This is in sharp contrast to the noise-removing algorithm used for the present purpose so far. We do not feel the need of complexity analysis of the present algorithm as it can easily run on a desktop or laptop of normal configurations.

The present day digital libraries have to use all modern network and server technologies in order to supply services of a high quality. For faster access and retrieval of a data or metadata smaller file size is one of the issues. The present technique can effectively be applied for the files supposed to serve the user requests.

References

Chowdhury, C G and Chowdhury S. 2003. *"Introduction to Digital Libraries."* Facet Publishing.

Wu, Y D, and Liu M. 2001. *"Content Management and the Future of Academic Libraries"*, The Electronic Library, Vol. 19(6), pp. 432-439.

Warren P and Alsmeyer D. 2005. "*The Digital Library: A Case Study in Intelligent Content Management*", *Journal of Knowledge Management*, Vol. **9**(5), pp. 28-39.

Nagy G. 2000. *"Twenty Years of Document Image Analysis"*, IEEE Trans. Pattern Analysis and Machine Intelligence, 22/1, pp. 38-62.

Chaudhuri B B, U Garain and M Mitra. "*On OCR of The Most Popular Indian Scripts: Devnagari and Bangla*", N. Murshed (Ed.), Singapore, World Scientific.

Ray Chaudhuri A, Mandal Achintya K, Chaudhuri B B. 2002. "*Page Layout Analyser for Multilingual Indian Documents*" Proc. of Language Engineering Conference-2002, IEEE CS Press, pp. 24-32.

Gonzalez R C, Woods R E. 2004. "*Digital Image Processing*", *Pearson Education, Delhi.*

Weeks Arthur R. 2007. "*Fundamentals of Electronic Image Processing*", Prentice Hall of India, New Delhi.

Holley R. 2010. "*Trove: Innovation in Access to Information in Australia*", issue 64, July 2010 (http://www.ariadne.ac.uk/issue64/holley/).

Holley Rose. 2011. "*Resource Sharing in Australia: Find and Get in Trove — Making 'Getting' Better*". D-Lib Magazine Vol 17, Number 3/4.