Optimal Document Archiving and Fast Information Retrieval

Hazem M. El-Bakry and Ahmed A. Mohammed

Abstract**—**In this paper, an intelligent algorithm for optimal document archiving is presented. It is kown that electronic archives are very important for information system management. Minimizing the size of the stored data in electronic archive is a main issue to reduce the physical storage area. Here, the effect of different types of Arabic fonts on electronic archives size is discussed. Simulation results show that PDF is the best file format for storage of the Arabic documents in electronic archive. Furthermore, fast information detection in a given PDF file is introduced. Such approach uses fast neural networks (FNNs) implemented in the frequency domain. The operation of these networks relies on performing cross correlation in the frequency domain rather than spatial one. It is proved mathematically and practically that the number of computation steps required for the presented FNNs is less than that needed by conventional neural networks (CNNs). Simulation results using MATLAB confirm the theoretical computations.

Keywords**—**Information Storage and Retrieval, Electronic Archiving, Fast Information Detection, Cross Correlation, Frequency Domain.

I. Introduction

CURRENTLY, the electronic archiving process of documents becomes indispensable. It introduces many advantages such as faster accessing to information and storage quality and other benefits [2,3,9]. The main objective of this paper is to develop and improve the work presented in [19].

The selecting of the best techniques and methods of data compression is a key concern in the electronic archiving process. Due to the huge volume of documents being preserved, huge storage area is required which reflected in the cost [1,3,1,11].

This work is an applied research discussing the impact of using different types of Arabic Fonts on the volume of data stored in electronic archiving systems [4,7,8]. This research depends on applying seven types of the most common Arab Fonts using four different Arab documents and studing the impact of each Font Type for each document. Five different types of widely used file formats, acrobat Format (PDF), Jpeg, GIFF,TIFF and Word Format, are used [3,10,16]. Also, all of the elements that might affect the results other than the Font type, such as: Font size, distance between the lines (line space), the content of the document, and paper size are fixed [14]. Moreover, an intelligent searching algorithm for fast word detection in a given file is presented.

A) Fonts

The text of a document must be shown visually and characters must be mapped to abstract glyphs. One or more characters can be depicted by one or more glyphs in a probably context dependent way. A glyph is the real artistic representation of an abstract glyph, in a certain typographical model, in the form of contours or bitmaps which can be drawn on the screen or paper. A font is a whole of glyphs, observing all the same basic reason according to the design, Size, aspect, and other attributes associated with the whole unit, and to trace characters to abstract glyphs [6,8,14]. The Fonts are specified according to the following characteristics:

1. Font family

A family of font is a group of fonts, conceived to be employed in association and exhibiting similarities in the design. A member of the family can be italic, another bold, and another condensed or using small caps [12,14].

2. Font style

The font style specifies whether the text is to be rendered using a normal, italic, or oblique face. The italic is a more cursive face of companion to the face normal, but not as cursive as for him to make a face of manuscript. Oblique is the tilted shape of the normal face, and is generally employed like face of companion with sans-serif [6,14].

3. Variant

Font Variant indicates if the text must be rendered using the normal glyphs for lowercase characters or using small-caps glyphs for lowercase characters. A particular font can contain only the normal, only the small-hats, or the two types of glyph; this property is employed to require a suitable font and, if the font contains the two alternatives, the suitable glyphs [6,8,14].

4. Weight of font

It refers to boldness or lightness glyphs employed to return the text, relative with other fonts in the same family of font [6,8,14].

5. Font stretch

H. M. El-Bakry is assistant professor with Dept. of Information Systems - Faculty of Computer Science and Information Systems – Mansoura University – Egypt. (phone: +2-050-2349340, fax: +2-050-2221442, e-mail: helbakry20@yahoo.com).

A. A. Mohammed is assistant professor with Dept. of Information Technology - Faculty of Computer Science and Information Systems – Mansoura University – Egypt. (phone: +2-050-2349340, fax: +2-050-2221442, e-mail: helbakry20@yahoo.com).

The font stretch indicates the desired quantity of condensation or expansion in the glyphs employed to return the text, relative to other fonts in the same family [4,12,14].

6. Font size

The size of a font refers to the size from baseline to baseline [12,14].

B) Archiving File Format

There are many formats and electronic technologies to select for archiving. Those include the ASCII (for the text), TIFF, PDF, JPEG, and XML-not to mention the text processing, the assessments, and other formats. The nature of patent rights of some of these formats leads to the criticism which they cannot be guaranteed to continue for the term. Only one of these formats is only agreed upon to ensure posting conservation over a long period [15,16].

The PDF represents not only the data contained in the document but also the exact form which the document took. The file can be viewed without original application. In fact, ten years as of now, and in the future, the users will be able to always view at the file exactly because it was created. With the addition of the metadata of XML to the PDF file, we can have fidelity and accessibility.

Since the PDF specification is publicly available, information about the file format is always available in the public domain, making it a very attractive format to choose for the electronic files. People with incapacities can also reach information by using assistive technology [10,13,15,16].

For example, a visually impaired person might use a screen reader, available from vendors such as Freedom Scientific, Dolphin Oceanic, and GW Micro, to verbalize the text. This is done through embedded tags in the PDF file structure. These tags can be created automatically from the originating application or entered as part of an enhancement process [10,13,15,16].

C) File Formats

There are three different categories of file formats: bitmap, vector, and metafiles. Information of a file stored as a bitmap is stored as a pattern of pixels. When Information of a file stored as a vector file, its information is stored as mathematical data. The metafile format can store an image's information as pixels, mathematical data, or both [15,16].

This Part briefly describes some of the most common formats which are usually selected for archiving purpose.

1. PDF format

In 1992, John Warnock, co-founder of Adobe Systems Incorporated, speaking about the goals of a development Project known as Camelot, said, The Camelot project developed the technology known as PDF. PDF leveraged the ability of the PostScript language to render complex text and graphics and brought this feature to the screen as well as the printer. Since then, the updated versions of the specifications of PDF continue to be provided by the adobe via the Web. The term *Portable Document Format*, or the PDF, was invented to illustrate that a file in accordance with this specifications can be viewed at and printed on any platform UNIX®, Mac OS, Microsoft® Windows®, and several mobile devices as well-with same fidelity. A document of PDF is the same one for the unspecified one among it platforms. It is composed of an order of the pages, with each page including/understanding the text, characteristic of font, margins, layout, graphical elements, and background and text colors. With all this wealth of information, the file of PDF can be reflected exactly on the screen and the printing device of. It can also include other items such as metadata, hyperlinks, and fields of form [13,15,16].

2. GIF Format

The Graphics Interchange Format (GIF) in the beginning was developed by CompuServe in 1987. It is one of the most popular formats of file for exchange of files graphs between the computers. It is most commonly used for bitmap images composed of line drawings or blocks of a few distinct colors. The format of GIF supports 8 bits of information or less color. Moreover, the format of GIF file supports transparency, allowing you to create a color in your transparent image [15,16].

3. JPEG format

The Joint Photographic Experts Group format (JPEG) is one of the most popular formats for web graphs. It supports 24 bits of information of color, and is generally used for photographs of the file Formats-2. The format of JPEG file stores all the information of color in an image of RGB, and then reduced the volume of file by compressing it, or by saving only the color information which is essential with the image. The majority of the applications let the user determine the quantity of compression used by saving a graph in the format of JPEG. With the difference of the GIF, the JPEG does not support transparency [10,15].

4. TIFF

The Tag Interchange File Format (TIFF) is a tag-based format that was developed and maintained by Aldus (now Adobe). TIF, which is used for bitmap images, is compatible with a wide range of software applications and can be used across platforms such as Macintosh, Windows, and UNIX. The TIFF format is complex, so TIFF files are generally larger than GIF or JPEG files [15].

II. Measuring the Effect of Arabic Fonts on File Size

This part of the paper is an applied study. It Studies the effect of using different types of Arabic Fonts on the size of data stored in electronic archive.

This research is applied for eight different types of the most common Arab Fonts (Times New Roman, Arial, Arabic Transparent, Diwani Letter, Simplified Arabic, Tahoma, Andalus, and Old Antic Bold). Four different Arab documents are used and the impact of each type of font is studied. In this research we used five different types of file format (PDF, JPEG, TIFF, GIF and Word format (Microsoft word)) which are more widely used.

Here we have fixed all of the elements that might affect the results other than the Font type, such as: Font size, distance between the lines (Line space), the content of the document, and paper size. Figure 1 shows examples for the used documents with different fonts. Table I, II, III and IV shows the experimental results for the documents no. 1, 2, 3 and 4 with the eight Arab fonts. Figures 2:10 represent the results in a graphical form

III. Fast Information Retreival from PDF Files by using Neural Networks and Cross Correlation in the Frequency Domain

First neural networks are trained to classify sub-matrices which contain the required information from those which do not and this is done in the spatial domain. In the test phase, each sub-matrix in the input file is tested for the presence or absence of the required information. At each position in the input file each sub-matrix is multiplied by a window of weights, which has the same size as the sub-matrix. This multiplication is done in the spatial domain. The outputs of neurons in the hidden layer are multiplied by the weights of the output layer. When the final output is high this means that the sub-matrix under test contain diseases and vice versa. Thus, we may conclude that this searching problem is cross correlation in the spatial domain between the file under test and the input weights of neural networks.

In this section, a fast algorithm for information detection based on two dimensional cross correlations that take place between the tested file and the sliding window (20x20 elements) is described. Such window is represented by the neural network weights situated between the input unit and the hidden layer. The convolution theorem in mathematical analysis says that a convolution of *f* with *h* is identical to the result of the following steps: let *F* and *H* be the results of the Fourier transformation of f and h in the frequency domain. Multiply *F* and *H* in the frequency domain point by point and then transform this product into spatial domain via the inverse Fourier transform [20-22]. As a result, these cross correlations can be represented by a product in the frequency domain. Thus, by using cross correlation in the frequency domain a speed up in an order of magnitude can be achieved during the detection process [25-58].

In the detection phase, a sub-matrix *X* of size *mxz* (sliding window) is extracted from the tested file, which has a size *P*x*T*, and fed to the neural network. Let W_i be the vector of weights between the input sub-matrix and the hidden layer. This vector has a size of *mxz* and can be represented as *mxz* matrix. The output of hidden neurons h_i can be calculated as follows [20-22]:

$$h_i = g\left(\sum_{j=1}^{m}\sum_{k=1}^{z} W_i(j,k)X(j,k) + b_i\right) \qquad (1)$$

where *g* is the activation function and b_i is the bias of each hidden neuron *(i)*. Eq. 1 represents the output of each hidden neuron for a particular sub-matrix *I*. It can be computed for the whole file Ψ as follows:

$$h_i(uv)=g\left(\sum_{j=-m/2}^{m/2}\sum_{k=-z/2}^{z/2} W_i(j,k)\,\Psi(u+j,v+k)+b_i\right) \qquad (2)$$

Eq.(2) represents a cross correlation operation. Given any two functions *f* and *g*, their cross correlation can be obtained by [23]:

$$g(x,y)\otimes f(x,y)=\left(\sum_{m=-\infty}^{\infty}\sum_{z=-\infty}^{\infty} g(m,z)f(x+m,y+z)\right) \qquad (3)$$

Therefore, Eq.(2) can be written as follows [20-22]:

$$h_i = g\left(W_i \otimes \Psi + b_i\right) \qquad (4)$$

where h_i is the output of the hidden neuron *(i)* and $h_i\,(u,v)$ is the activity of the hidden unit *(i)* when the sliding window is located at position *(u,v)* in the input file Ψ and $(u,v)\in[P-m+1,T-n+1]$.

Now, the above cross correlation can be expressed in terms of the Fourier Transform [20-22]:

$$W_i \otimes \Psi = F^{-1}\left(F(\Psi)\bullet F^*\left(W_i\right)\right) \qquad (5)$$

(*) means the conjugate of the *FFT* for the weight matrix. Hence, by evaluating this cross correlation, a speed up ratio can be obtained comparable to conventional neural networks. Also, the final output of the neural network can be evaluated as follows [20-22]:

$$O(u,v) = g\left(\sum_{i=1}^{q} W_o(i)\,h_i(u,v) + b_o\right) \qquad (6)$$

where *q* is the number of neurons in the hidden layer. *O(u,v)* is the output of the neural network when the sliding window located at the position *(u,v)* in the input file Ψ. W_o is the weight matrix between hidden and output layer.

The complexity of cross correlation in the frequency domain can be analyzed as follows:

1. For a tested file of *NxN* elements, the *2D-FFT* requires a number equal to $N^2log_2N^2$ of complex computation steps. Also, the same number of complex computation steps is required for computing the *2D-FFT* of the weight matrix for each neuron in the hidden layer.

2. At each neuron in the hidden layer, the inverse *2D-FFT* is computed. So, *q* backward and *(1+q)* forward transforms have to be computed. Therefore, for a file under test, the total number of the *2D-FFT* to compute is $(2q+1)N^2log_2N^2$.

3. The input file and the weights should be multiplied in the frequency domain. Therefore, a number of complex computation steps equal to qN^2 should be added.

4. The number of computation steps required by the fast neural networks is complex and must be converted into a real version. It is known that the two dimensional Fast Fourier Transform requires $(N^2/2)log_2N^2$ complex multiplications and $N^2log_2N^2$ complex additions [23]. Every complex multiplication is realized by six real floating point operations and every complex addition is implemented by two real floating point operations. So, the total number of computation steps required to obtain the *2D-FFT* of an *NxN* file is:

$$\rho=6((N^2/2)log_2N^2) + 2(N^2log_2N^2) \qquad (7)$$

which may be simplified to:

$$\rho=5N^2log_2N^2 \qquad (8)$$

Performing complex dot product in the frequency domain also requires $6qN^2$ real operations.

5. In order to perform cross correlation in the frequency domain, the weight matrix must have the same size as the input file. Assume that the input object has a size of (nxn) dimensions. So, the search process will be done over sub-matrix of (nxn) dimensions and the weight matrix will have the same size. Therefore, a number of zeros = (N^2-n^2) must be added to the weight matrix. This requires a total real number of computation steps = $q(N^2-n^2)$ for all neurons. Moreover, after computing the *2D-FFT* for the weight matrix, the conjugate of this matrix must be obtained. So, a real number of computation steps $=qN^2$ should be added in order to obtain the conjugate of the weight matrix for all neurons. Also, a number of real computation steps equal to *N* is required to create butterflies complex numbers $(e^{-jk(2\Pi n/N)})$, where $0<K<L$. These *(N/2)* complex numbers are multiplied by the elements of the input file or by previous complex numbers during the computation of the *2D-FFT*. To create a complex number requires two real floating point operations. So, the total number of computation steps required for the fast neural networks becomes:

$$\sigma=(2q+1)(5N^2log_2N^2) +6qN^2+q(N^2-n^2)+qN^2 +N \qquad (9)$$

which can be reformulated as:

$$\sigma=(2q+1)(5N^2log_2N^2) +q(8N^2-n^2) +N \qquad (10)$$

6. Using a sliding window of size nxn for the same file of *NxN* elements, $q(2n^2-1)(N-n+1)^2$ computation steps are required when using traditional neural networks for object detection process. The theoretical speed up factor η can be evaluated as follows:

$$\eta=\frac{q(2n^2-1)(N-n+1)^2}{(2q+1)(5N^2log_2N^2)+q(8N^2-n^2)+N} \qquad (11)$$

The theoretical speed up ratio Eq. 11 with different sizes of the input file and different in size weight matrices is listed in Table V. Practical speed up ratio for manipulating files of different sizes and different in size weight matrices is listed in Table VI using 2.7 GHz processor and *MATLAB ver 5.3*. An interesting property with FNNs is that the number of computation steps does not depend on either the size of the input sub-matrix or the size of the weight matrix (n). The effect of (n) on the number of computation steps is very small and can be ignored. This is in contrast to CNNs in which the number of computation steps is increased with the size of both the input sub-matrix and the weight matrix (n).

V. CONCLUSION

A new fast algorithm for information retrieval from a given file has been presented. This has been achieved by performing cross correlation in the frequency domain between input file and the input weights of fast neural networks (FNNs). It has been proved mathematically and practically that the number of computation steps required for the presented FNNs is less than that needed by conventional neural networks (CNNs). Simulation results using MATLAB has confirmed the theoretical computations. In addition, this paper studied the effect of using different types of Arabic Fonts on the volume of data stored in an electronic archive including Arabic documents. Seven types of Fonts were used in four different Arab documents with five different types of file format. The result indicated that:

1. The Arabic document's size which is stored in an electronic archive changes according to the used Arabic font type.
2. Some Arabic Fonts (Diwani, Arabic transparence, Simplify Arabic) result in reduced file size than others fonts.
3. Some Arabic Fonts(for example Tohama) are not suitable for archiving the Arabic documents
4. Some non famous fonts (Diwani font) give the best results than the common fonts used (Arabic transparence, Simplify Arabic).
5. Arabic transparence and Simplify Arabic fonts are suitable for PDF Format.
6. The PDF is the best file format for storage of the Arabic documents in electronic archive.
7. The TIFF format is not suitable for archiving Arabic documented.
8. The word format is not affected by font type.

REFERENCES

[1] Blake, Monica," Archiving of Electronic Publications," Electronic Library, v7 n6 p376-86 Dec 1989
[2] Blake, Monica," Aspects of Electronic Archives," Electronic Publishing Review, v6 n3 p151-67 Sep 1986
[3] GAIL M., HODGE, "Best Practices for Digital Archiving," January 2000 issue of D-Lib Magazine, Volume 6 Number 1
[4] Ibrahim S. I. "Arabic Font Recognition using Decision Trees Built from Common Words," Journal of Computing and Information Technology - CIT 13, 2005, 3, 211–22.
[5] James A. Storer, Thomas G., "Data compression via textual substitution," Journal of the ACM (JACM) Volume 29, Issue 4, October 1982, Pages: 928 – 951.
[6] Jan A., Bernard J., "Font compression and retrieval," US Patent Issued on May 1, 2007.
[7] Khorsheed M.S., Clocksin, W.F., Spectral features for Arabic word recognition," Acoustics, Speech, and Signal Processing, ICASSP, Proceedings IEEE International Conference, 2000.
[8] Mohammad S., William F.," Multi-Font Arabic Word Recognition Using Spectral Features", 15th International Conference on Pattern Recognition (ICPR'00) - Volume 4 p. 4543
[9] Namane A., Sid-Ahmed M.A. ," Character scaling by contour method", Pattern Analysis and Machine Intelligence, IEEE Transactions on, Jun 1990, Volume:12, page(s):600-606.
[10] Sayood K., "Introduction to Data Compression," Morgan Kaufmann, 2006
[11] Syed A. A.," System, method and computer program product for generic outline font compression," United States Patent 6,614,940, September 2, 2003.
[12] Thomas A. Phelps, Robert W.," Two diet plans for fat PDF," Proceedings of the ACM symposium on Document engineering Grenoble, France, 2003, Pages: 175 – 184.
[13] "Fonts", http://www.w3.org/TR/REC-CSS2/fonts.html

[14] "Graphic File Formats at a Glance", www.visl.technion.ac.il/labs/anat/2/fileformats.pdf

[15] "PDF as a Standard for Archiving", http://www.adobe.com/enterprise/pdfs/pdfarchiving.pdf

[16] "U.K. Records Management for Central Government", http://*www.pro.gov.uk/recordsmanagement*

[17] "Victorian Electronic Records Strategy Standards and Guides", http://*www.prov.vic.gov.au/vers/standards/standards.htm*

[18] "The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project", http://*www.interpares.org/book/index.cfm*

[19] A. A. Mohammed ," The Effect of Arabic Fonts on Electronic Archive Size", Mansoura Journal for computer science and information systems, vol. 4, No. 4, 2007

[20] H. M. El-Bakry, "A Novel High Speed Delay Neural Model for Fast Pattern Recognition," Accepted for publication in Soft Computing Journal.

[21] H. M. El-Bakry, "Fast Virus Detection by using High Speed Time Delay Neural Networks," Accepted for publication in journal of computer virology.

[22] H. M. El-Bakry, "New Fast Principal Component Analysis For Real-Time Face Detection," Accepted for publication in MG&V Journal.

[23 J. W. Cooley, and J. W. Tukey, An algorithm for the machine calculation of complex Fourier series, Math. Comput. 19, 297–301 1965.

[24] Klette R., and Zamperon, "Handbook of image processing operators, " John Wiley & Sonsltd, 1996.

[25] H. M. El-bakry, "An Efficient Algorithm for Pattern Detection using Combined Classifiers and Data Fusion," Accepted for publication in Information Fusion Journal.

[26] Hazem M. El-bakry, and Mohamed Hamada "Fast Time Delay Neural Networks for Detecting DNA Coding Regions," Proc. of Kes 2009, Part I, LNAI 5711, Springer, September 28-30, 2009, pp. 334-342.

[27] H. M. El-Bakry and M. Hamada, " New Fast Decision Tree Classifier for Identifying Protein Coding Regions," Lecture Notes in Computer Science, Springer, ISICA 2008, LNCS 5370, 2008, pp. 489-500.

[28] H. M. El-Bakry and M. Hamada, "A New Implementation for High Speed Neural Networks in Frequency Space," Lecture Notes in Artificial Intelligence, Springer, KES 2008, Part I, LNAI 5177, pp. 33-40.

[29] H. M. El-Bakry, "New Faster Normalized Neural Networks for Sub-Matrix Detection using Cross Correlation in the Frequency Domain and Matrix Decomposition," Applied Soft Computing journal, vol. 8, issue 2, March 2008, pp. 1131-1149.

[30] H. M. El-Bakry, and Nikos Mastorakis "New Fast Normalized Neural Networks for Pattern Detection," Image and Vision Computing Journal, vol. 25, issue 11, 2007, pp. 1767-1784.

[31] H. M. El-Bakry and Nikos Mastorakis, "Fast Code Detection Using High Speed Time Delay Neural Networks," Lecture Notes in Computer Science, Springer, vol. 4493, Part III, May 2007, pp. 764-773.

[32] H. M. El-Bakry, "New Fast Principal Component Analysis for Face Detection," Journal of Advanced Computational Intelligence and Intelligent Informatics, vol.11, no.2, 2007, pp. 195-201.

[33] H. M. El-Bakry, "New Fast Time Delay Neural Networks Using Cross Correlation Performed in the Frequency Domain," Neurocomputing Journal, vol. 69, October 2006, pp. 2360-2363.

[34] H. M. El-Bakry, and Nikos Mastorakis, "A Novel Model of Neural Networks for Fast Data Detection," WSEAS Transactions on Computers, Issue 8, vol. 5, November 2006, pp. 1773-1780.

[35] H. M. El-Bakry, and N. Mastorakis, "A New Approach for Fast Face Detection," WSEAS Transactions on Information Science and Applications, issue 9, vol. 3, September 2006, pp. 1725-1730.

[36] H. M. El-Bakry, "A New Implementation of PCA for Fast Face Detection," International Journal of Intelligent Technology, Vol. 1, No.2, 2006, pp. 145-153.

[37] H. M. El-Bakry, and Q. Zhao, "Fast Normalized Neural Processors For Pattern Detection Based on Cross Correlation Implemented in the Frequency Domain," Journal of Research and Practice in Information Technology, Vol. 38, No.2, May 2006, pp. 151-170.

[38] H. M. El-Bakry, "Fast Painting with Different Colors Using Cross Correlation in the Frequency Domain," International Journal of Computer Science, vol.1, no.2, 2006, pp. 145-156.

[39] H. M. El-Bakry, "Faster PCA for Face Detection Using Cross Correlation in the Frequency Domain," International Journal of Computer Science and Network Security, vol.6, no. 2A, February 2006, pp.69-74.

[40] H. M. El-Bakry, "New High Speed Normalized Neural Networks for Fast Pattern Discovery on Web Pages," International Journal of Computer Science and Network Security, vol.6, No. 2A, February 2006, pp.142-152.

[41] H. M. El-Bakry, and Q. Zhao, "Fast Time Delay Neural Networks," International Journal of Neural Systems, vol. 15, no.6, December 2005, pp.445-455.

[42] H. M. El-Bakry, and Q. Zhao, "Speeding-up Normalized Neural Networks For Face/Object Detection," Machine Graphics & Vision Journal (MG&V), vol. 14, No.1, 2005, pp. 29-59.

[43] H. M. El-Bakry, "Pattern Detection Using Fast Normalized Neural Networks," Lecture Notes in Computer Science, Springer, vol. 3696, September 2005, pp. 447-454.

[44] H. M. El-Bakry, "Human Face Detection Using New High Speed Modular Neural Networks," Lecture Notes in Computer Science, Springer, vol. 3696, September 2005, pp. 543-550.

[45] H. M. El-Bakry, and Q. Zhao, "Fast Pattern Detection Using Normalized Neural Networks and Cross Correlation in the Frequency Domain," EURASIP Journal on Applied Signal Processing, Special Issue on Advances in Intelligent Vision Systems: Methods and Applications—Part I, vol. 2005, no. 13, 1 August 2005, pp. 2054-2060.

[46] H. M. El-Bakry, "A New High Speed Neural Model For Character Recognition Using Cross Correlation and Matrix Decomposition," International Journal of Signal Processing, vol.2, no.3, 2005, pp. 183-202.

[47] H. M. El-Bakry, and Q. Zhao, "A Fast Neural Algorithm for Serial Code Detection in a Stream of Sequential Data," International Journal of Information Technology, vol.2, no.1, pp. 71-90, 2005.

[48] H. M. El-Bakry, and Q. Zhao, "Fast Complex Valued Time Delay Neural Networks," International Journal of Computational Intelligence, vol.2, no.1, pp. 16-26, 2005.

[49] H. M. El-Bakry, and Q. Zhao, "Fast Pattern Detection Using Neural Networks Realized in Frequency Domain," Enformatika Transactions on Engineering, Computing, and Technology, February 25-27, 2005, pp. 89-92.

[50] H. M. El-Bakry, and Q. Zhao, "Sub-Image Detection Using Fast Neural Processors and Image Decomposition," Enformatika Transactions on Engineering, Computing, and Technology, February 25-27, 2005, pp. 85-88.

[51] H. M. El-Bakry, and Q. Zhao, "Face Detection Using Fast Neural Processors and Image Decomposition," International Journal of Computational Intelligence, vol.1, no.4, 2004, pp. 313-316.

[52] H. M. El-Bakry, and Q. Zhao, "A Modified Cross Correlation in the Frequency Domain for Fast Pattern Detection Using Neural Networks," International Journal on Signal Processing, vol.1, no.3, 2004, pp. 188-194.

[53] H. M. El-Bakry, and Q. Zhao, "Fast Object/Face Detection Using Neural Networks and Fast Fourier Transform," International Journal on Signal Processing, vol.1, no.3, 2004, pp. 182-187.

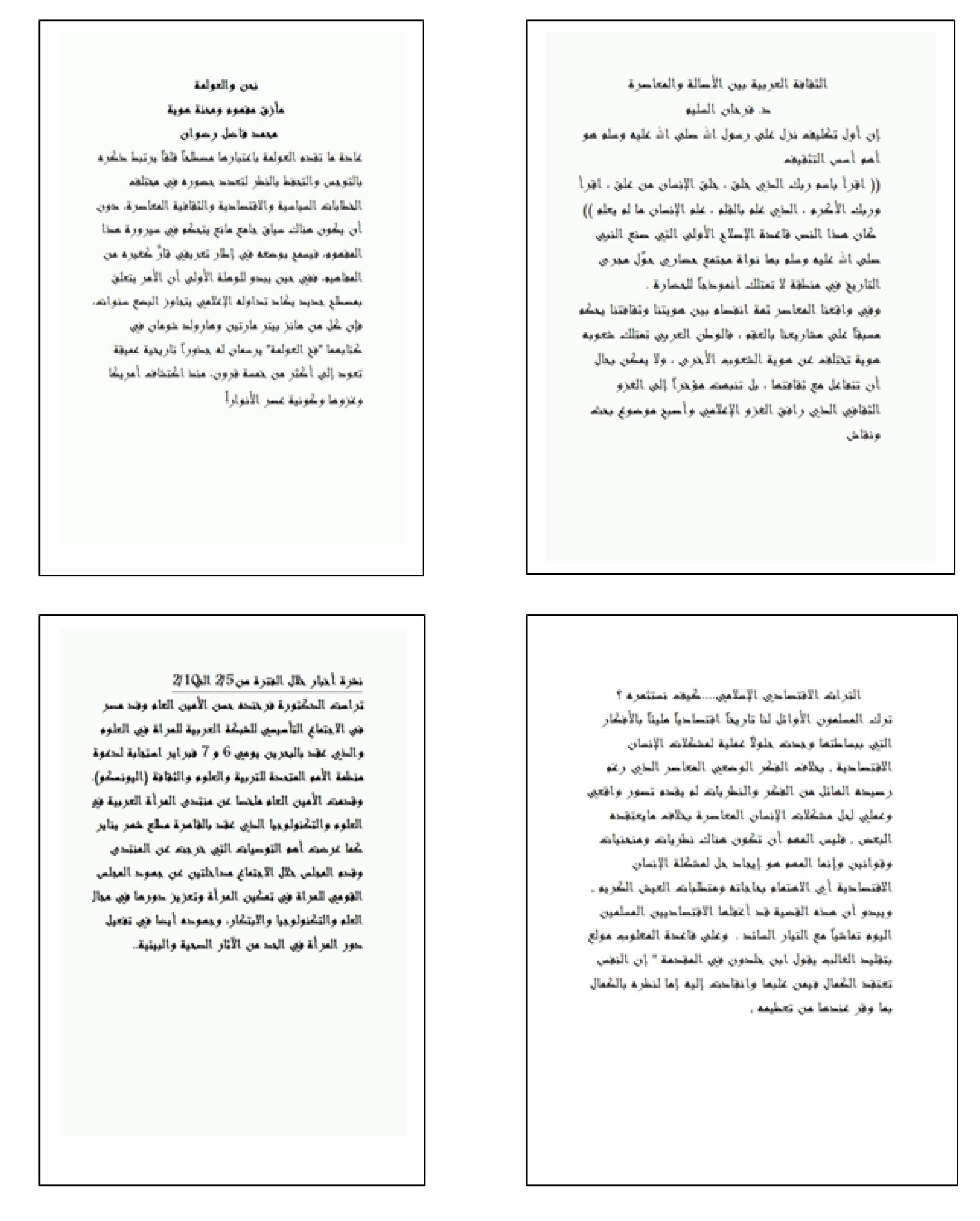
[54] H. M. El-Bakry, "Face detection using fast neural networks and image decomposition," Neurocomputing Journal, vol. 48, October 2002, pp. 1039-1046.

[55] H. M. El-Bakry, "Human Iris Detection Using Fast Cooperative Modular Neural Nets and Image Decomposition," Machine Graphics & Vision Journal (MG&V), vol. 11, no. 4, 2002, pp. 498-512.

[56] H. M. El-Bakry, "Fast Face Detection Using Neural Networks and Image Decomposition," Lecture Notes in Computer Science, Springer, vol. 2252, December, 2001, pp.205-215.

[57] H. M. El-Bakry "Fast Iris Detection for Personal Verification Using Modular Neural Networks," Lecture Notes in Computer Science, Springer, vol. 2206, October 2001, pp. 269-283.

[58] H. M. El-Bakry, "Automatic Human Face Recognition Using Modular Neural Networks," Machine Graphics & Vision Journal (MG&V), vol. 10, no. 1, 2001, pp. 47-73.



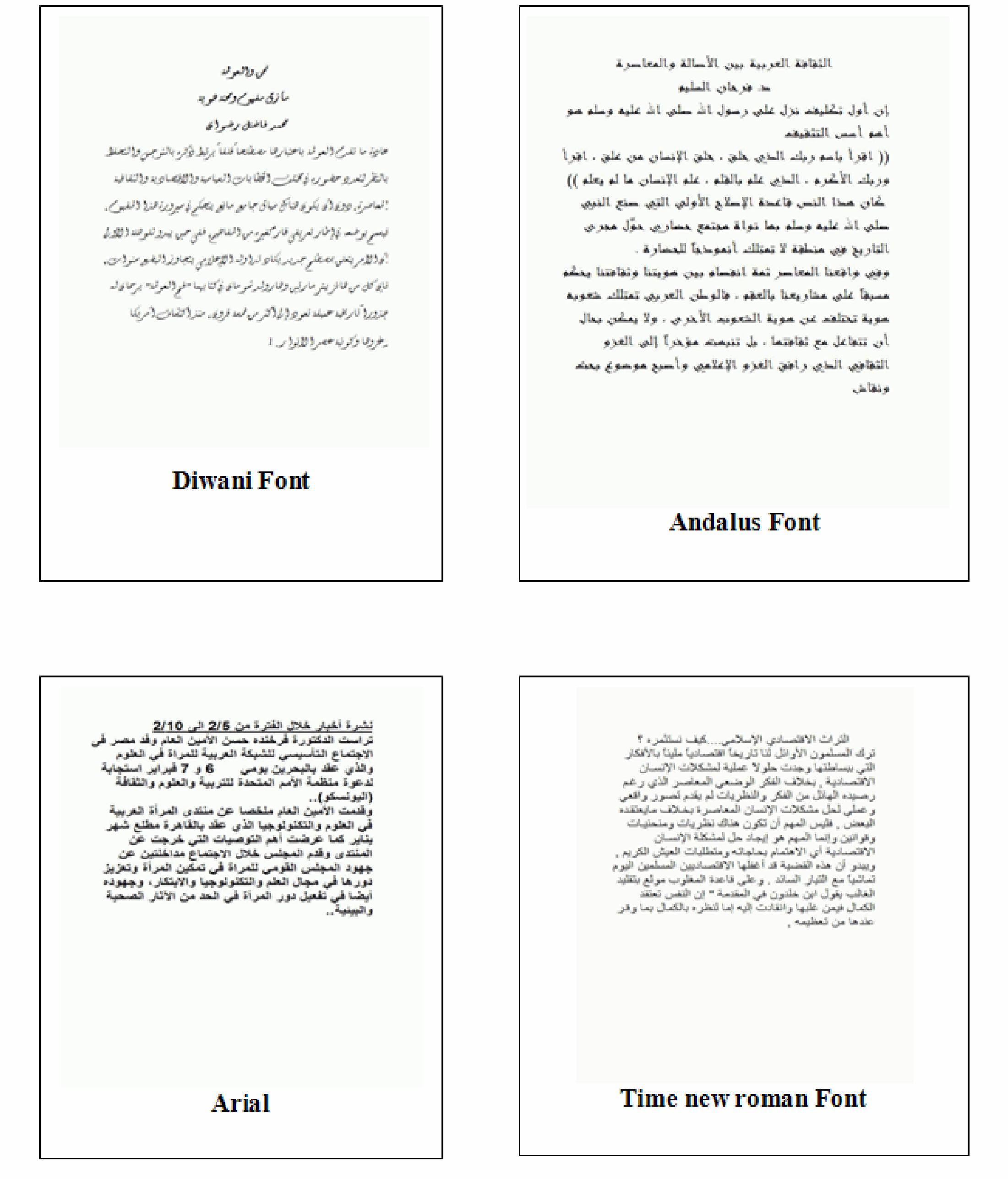
نحن والعولمة
مأزق مفهوم ومحنة هوية
محمد فاضل رضوان
عادة ما تقدم العولمة باعتبارها مصطلحاً قلقاً يرتبط ذكره
بالتوجس والتحفظ بالنظر لتعدد حضوره في مختلف
الخطابات السياسية والاقتصادية والثقافية المعاصرة، دون
أن يكون هناك سياق جامع مانع يتحكم في سيرورة هذا
المفهوم، فيسمح بوضعه في إطار تعريفي قارّ كغيره من
المفاهيم. ففي حين يبدو للوهلة الأولى أن الأمر يتعلق
بمصطلح جديد يكاد تداوله الإعلامي يتجاوز البضع سنوات،
فإن كلاً من هانز بيتر مارتين وهارولد شومان في
كتابهما "فخ العولمة" يرسمان له جذوراً تاريخية عميقة
تعود إلى أكثر من خمسة قرون، منذ اكتشاف أمريكا
وغزوها وكونية عصر الأنوار.

الثقافة العربية بين الأصالة والمعاصرة
د. فرحان السليم
إن أول تكليف نزل على رسول الله صلى الله عليه وسلم هو
أهم أسس التثقيف
((اقرأ باسم ربك الذي خلق . خلق الإنسان من علق . اقرأ
وربك الأكرم . الذي علم بالقلم . علم الإنسان ما لم يعلم))
كان هذا النص قاعدة الإصلاح الأولى التي صنع النبي
صلى الله عليه وسلم بها نواة مجتمع حضاري حوّل مجرى
التاريخ في منطقة لا تمتلك أنموذجاً للحضارة .
وفي واقعنا المعاصر ثمة انفصام بين هويتنا وثقافتنا يحكم
مسبقاً على مشاريعنا بالعقم . فالوطن العربي تمتلك شعوبه
هوية تختلف عن هوية الشعوب الأخرى . ولا يمكن بحال
أن تتفاعل مع ثقافتها . بل تنبهت مؤخراً إلى الغزو
الثقافي الذي رافق الغزو الإعلامي وأصبح موضوع بحث
ونقاش

نشرة أخبار خلال الفترة من2/5 إلى2/1
ترأست الدكتورة فرخندة حسن الأمين العام وفد مصر
في الاجتماع التأسيسي للشبكة العربية للمرأة في العلوم
والذي عقد بالبحرين يومي 6 و 7 فبراير استجابة لدعوة
منظمة الأمم المتحدة للتربية والعلوم والثقافة (اليونسكو).
وقدمت الأمين العام ملخصا عن منتدى المرأة العربية في
العلوم والتكنولوجيا الذي عقد بالقاهرة مطلع شهر يناير
كما عرضت أهم التوصيات التي خرجت عن المنتدى
وقدم المجلس خلال الاجتماع مداخلتين عن جهود المجلس
القومي للمرأة في تمكين المرأة وتعزيز دورها في مجال
العلم والتكنولوجيا والابتكار، وجهوده أيضا في تفعيل
دور المرأة في الحد من الآثار الصحية والبيئية.

التراث الاقتصادي الإسلامي....كيف نستثمره ؟
ترك المسلمون الأوائل لنا تاريخاً اقتصادياً مليئاً بالأفكار
التي بواسطتها وجدت حلولاً عملية لمشكلات الإنسان
الاقتصادية . بخلاف الفكر الوضعي المعاصر الذي رغم
رصيده الهائل من الفكر والنظريات لم يقدم تصور واقعي
وعملي لحل مشكلات الإنسان المعاصرة بخلاف مايعتقده
البعض . فليس المهم أن تكون هناك نظريات ومنحنيات
وقوانين وإنما المهم هو إيجاد حل لمشكلة الإنسان
الاقتصادية أي الاهتمام بحاجاته ومتطلبات العيش الكريم .
ويبدو أن هذه القضية قد أغفلها الاقتصاديين المسلمين
اليوم تماشياً مع التيار السائد . وعلى قاعدة المغلوب مولع
بتقليد الغالب يقول ابن خلدون في المقدمة " إن النفس
تعتقد الكمال فيمن غلبها وانقادت إليه إما لنظره بالكمال
بما وقر عندها من تعظيمه .

Fig. 1 Four various Arabic documents



نحن والعولمة
مازن سليم ومحمد عويد
محمد فاضل رضوان
[illegible]

Diwani Font

الثقافة العربية بين الأصالة والمعاصرة
د. فرحان السليم
إن أول تكليف نزل على رسول الله صلى الله عليه وسلم هو
أهم أسس التثقيف
((اقرأ باسم ربك الذي خلق . خلق الإنسان من علق . اقرأ
وربك الأكرم . الذي علم بالقلم . علم الإنسان ما لم يعلم))
كان هذا النص قاعدة الإصلاح الأولى التي صنع النبي
صلى الله عليه وسلم بها نواة مجتمع حضاري حوّل مجرى
التاريخ في منطقة لا تمتلك أنموذجاً للحضارة .
وفي واقعنا المعاصر ثمة انفصام بين هويتنا وثقافتنا يحكم
مسبقاً على مشاريعنا بالعقم . فالوطن العربي تمتلك شعوبه
هوية تختلف عن هوية الشعوب الأخرى . ولا يمكن بحال
أن تتفاعل مع ثقافتنا . بل تتعرض مؤخراً إلى الغزو
الثقافي الذي رافق الغزو الإعلامي وأصبح موضوع بحث
ونقاش

Andalus Font

نشرة أخبار خلال الفترة من 2/5 الى 2/10
ترأست الدكتورة فرخنده حسن الأمين العام وفد مصر في
الاجتماع التأسيسي للشبكة العربية للمرأة في العلوم
والذي عقد بالبحرين يومي 6 و 7 فبراير استجابة
لدعوة منظمة الأمم المتحدة للتربية والعلوم والثقافة
(اليونسكو)..
وقدمت الأمين العام ملخصا عن منتدى المرأة العربية
في العلوم والتكنولوجيا الذي عقد بالقاهرة مطلع شهر
يناير كما عرضت أهم التوصيات التي خرجت عن
المنتدى وقدم المجلس خلال الاجتماع مداخلتين عن
جهود المجلس القومي للمرأة في تمكين المرأة وتعزيز
دورها في مجال العلم والتكنولوجيا والابتكار، وجهوده
أيضا في تفعيل دور المرأة في الحد من الآثار الصحية
والبيئية..

Arial

التراث الاقتصادي الإسلامي....كيف نستثمره ؟
ترك المسلمون الأوائل لنا تاريخا اقتصاديا مليئا بالأفكار
التي ببساطتها وجدت حلولا عملية لمشكلات الإنسان
الاقتصادية , بخلاف الفكر الوضعي المعاصر الذي رغم
رصيده الهائل من الفكر والنظريات لم يقدم تصور واقعي
وعملي لحل مشكلات الإنسان المعاصرة بخلاف مايعتقده
البعض . فليس المهم أن تكون هناك نظريات ومنحنيات
وقوانين وإنما المهم هو إيجاد حل لمشكلة الإنسان
الاقتصادية أي الاهتمام بحاجاته ومتطلبات العيش الكريم .
ويبدو أن هذه القضية قد أغفلها الاقتصاديين المسلمين اليوم
تماشيا مع التيار السائد . وعلى قاعدة المغلوب مولع بتقليد
الغالب يقول ابن خلدون في المقدمة " إن النفس تعتقد
الكمال فيمن غلبها وانقادت إليه إما لنظره بالكمال بما وقر
عندها من تعظيمه .

Time new roman Font

Fig. 2 Examples for Arabic Fonts

TABLE I
THE REQUIRED STORAGE AMOUNT(IN KILO BYTES) FOR DOCUMENT NO.1

Font	WORD Format	JBEG Format	GIF Format	TIFF Format	PDF Format
Arial	24.5	67.6	54.5	313.4	35
Time New Roman	24.5	67.9	53.8	317.6	36.91
Diwani Letter	24.5	59.5	48	275.8	19.55
Arabic Transparence	24.5	67.6	54.5	315.2	20.87
Andalus	24.5	67.3	59.7	320.6	17.04
Old Antique Bold	24.5	65.4	61.6	325.6	18.54
Simplified Arabic	24.5	69	54.5	322	20.85
Tohama	24.5	90.1	74.3	429.4	31.01

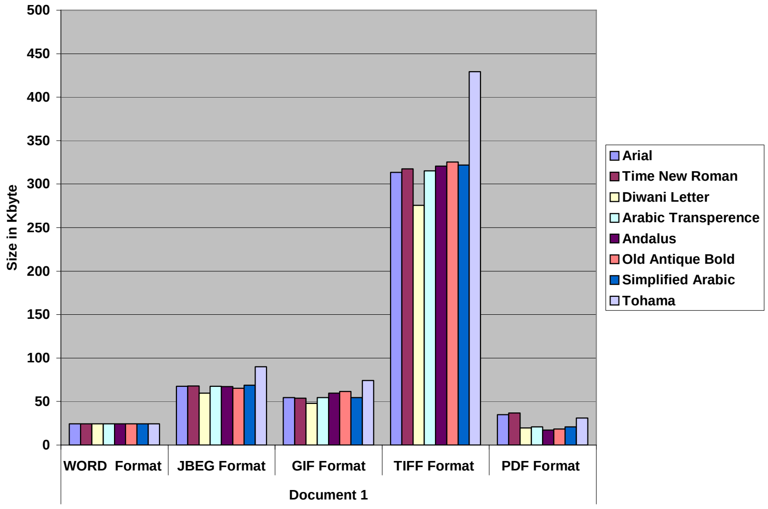


Fig. 3 File size at various Font style (for document No.1)

TABLE II
THE REQUIRED STORAGE AMOUNT(IN KILO BYTES) FOR DOCUMENT NO.2

Font	WORD Format	JBEG Format	GIF Format	TIFF Format	PDF Format
Arial	26.5	72.3	59	344.8	21.5
Time New Roman	26.5	72	58.8	337	24.29
Diwani Letter	26.5	66.4	55.4	311.5	19.35
Arabic Transparence	26.5	71.3	56.2	338	14.43
Andalus	26.5	71	59.5	342.2	16.33
Old Antique Bold	26.5	69.4	66.7	354.5	17.74
Simplified Arabic	26.5	72.5	58.6	350.1	14.45
Tohama	26.5	93.4	78.8	458.5	21.17

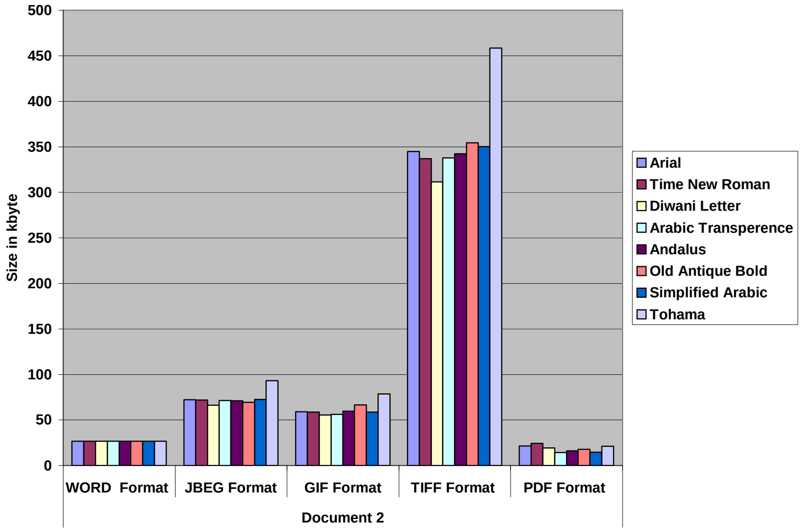


Fig. 4 File size at various Font style (for document No.2)

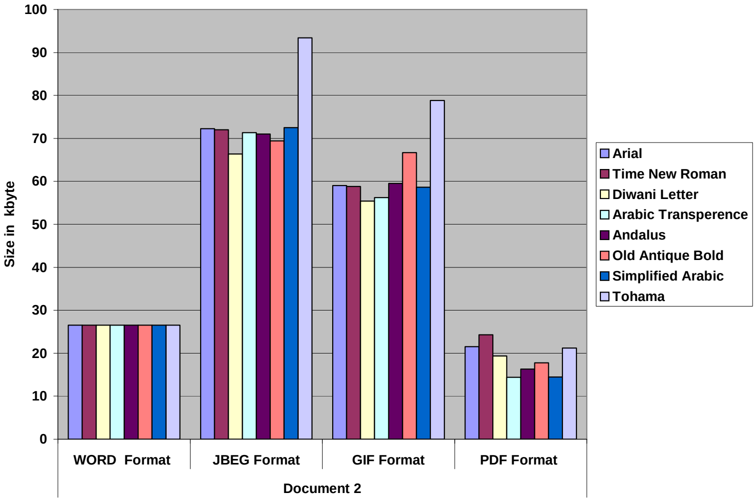


Fig. 5 File size at various Font style (for document No.2)

TABLE III
THE REQUIRED STORAGE AMOUNT(IN KILO BYTES) FOR DOCUMENT NO.3

Font	WORD Format	JBEG Format	GIF Format	TIFF Format	PDF Format
Arial	24.5	77	65.8	362.8	23.71
Time New Roman	24.5	77.6	63.8	362.7	24.25
Diwani Letter	24.5	68.7	61.1	319.1	18.7
Arabic Transparence	24.5	74.8	62.7	357.2	13.81
Andalus	24.5	74.1	59.7	350.9	16.29
Old Antique Bold	24.5	71.3	64.8	351.7	17.99
Simplified Arabic	24.5	76.9	66.6	371.1	13.8
Tohama	24.5	103.7	79.4	501.6	19.11

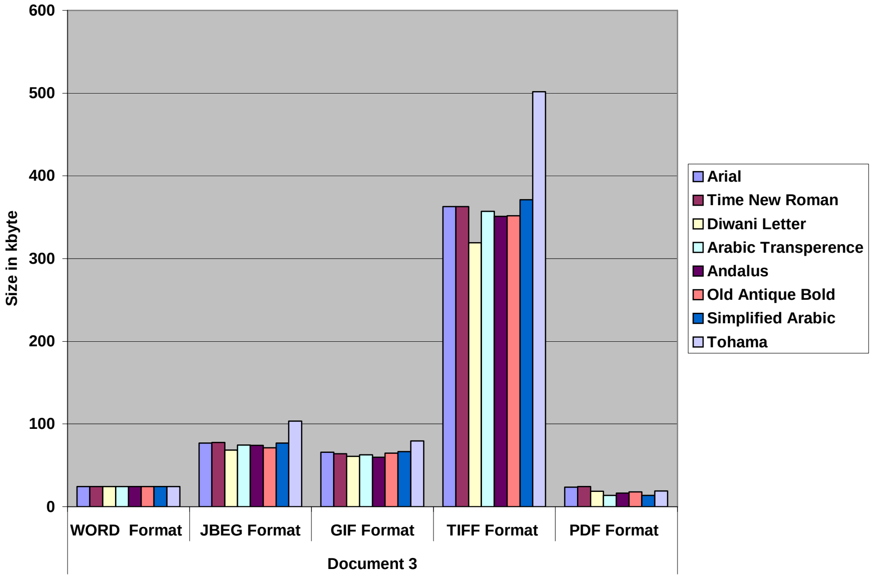


Fig. 6 File size at various Font style (for document No.3)

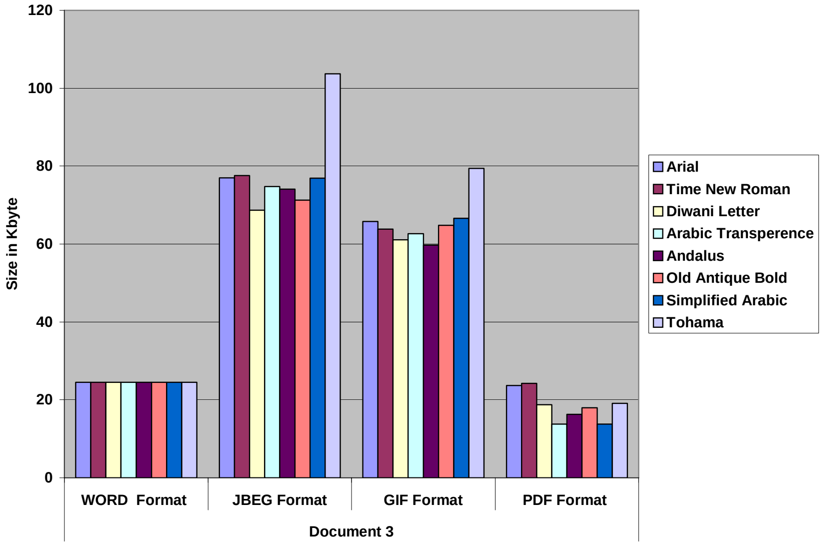


Fig. 7 File size at various Font style (for document No.3)

TABLE IV
THE REQUIRED STORAGE AMOUNT(IN KILO BYTES) FOR DOCUMENT NO. 4

Font	WORD Format	JBEG Format	GIF Format	TIFF Format	PDF Format
Arial	24.5	74.8	60.9	353.8	19.96
Time New Roman	24.5	74.5	61.9	355.5	21.14
Diwani Letter	24.5	69.4	60.8	324.5	18.58
Arabic Transparence	24.5	73.8	60.6	349.8	13.36
Andalus	24.5	74.6	66.9	360.9	15.25
Old Antique Bold	24.5	74.1	70.3	371.9	17.15
Simplified Arabic	24.5	74.6	61.3	358.6	13.37
Tohama	24.5	100.3	86.4	492.4	19.68

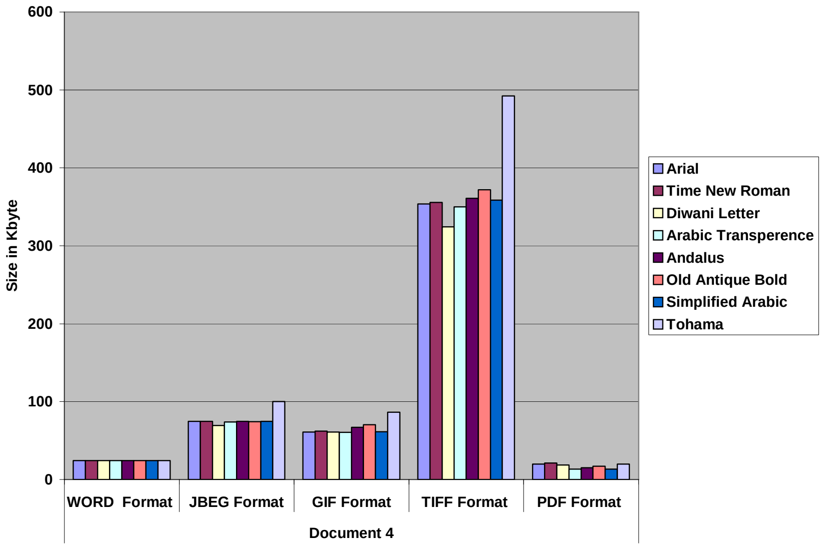


Fig. 8 File size at various Font style (for document No.4)

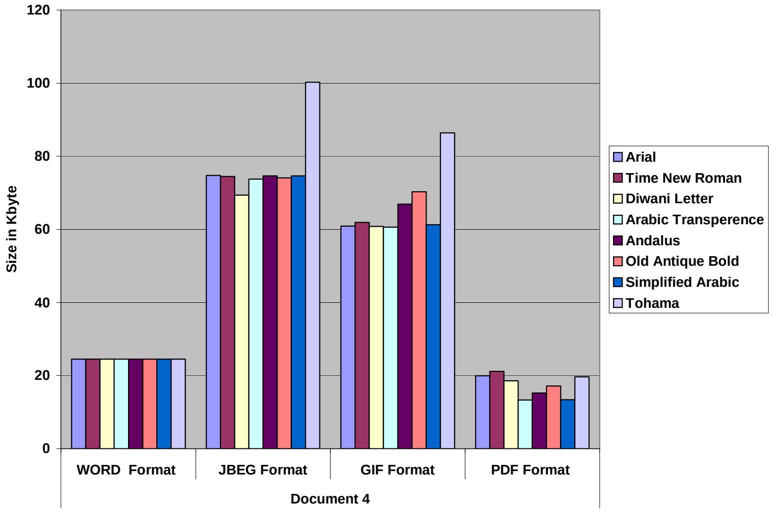


Fig. 9 File size at various Font style (for document No.4)

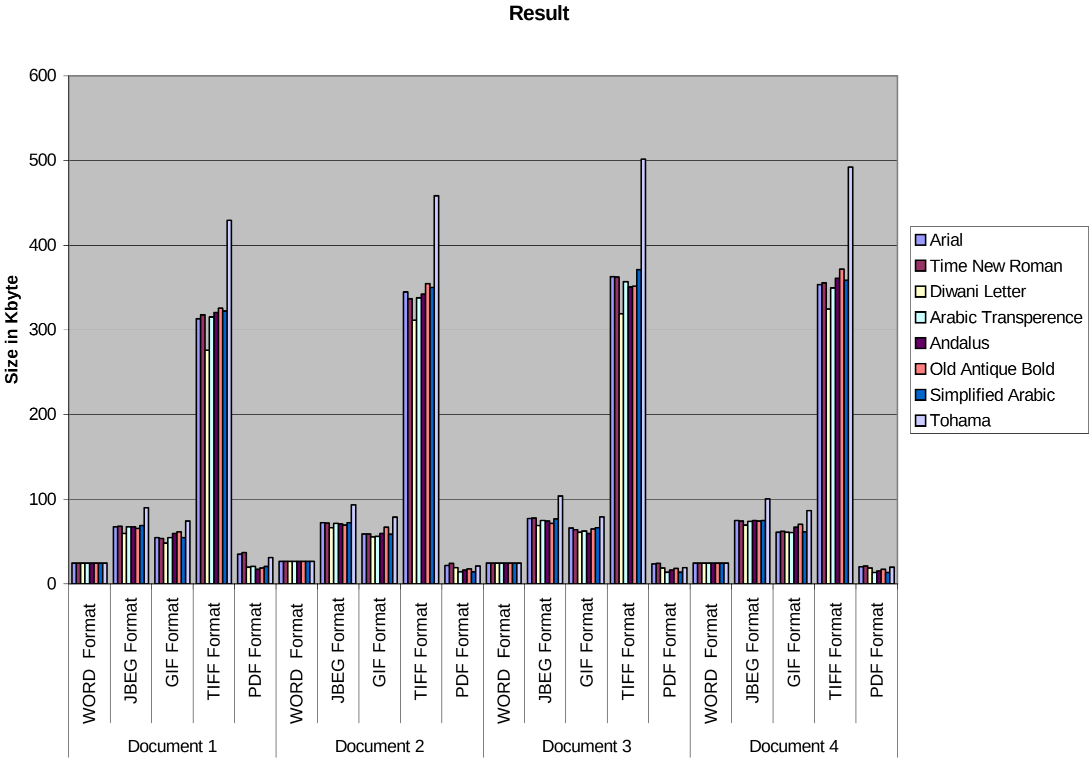


Fig. 10 File size at various Font style (for all documents)

TABLE V
THE THEORETICAL SPEED UP RATIO FOR INFORMATION RETREIVAL FROM FILES WITH DIFFERENT SIZES

Files size	Speed up ratio (n=20)	Speed up ratio (n=25)	Speed up ratio (n=30)
100x100	3.67	5.04	6.34
200x200	4.01	5.92	8.05
300x300	4.00	6.03	8.37
400x400	3.95	6.01	8.42
500x500	3.89	5.95	8.39
600x600	3.83	5.88	8.33
700x700	3.78	5.82	8.26
800x800	3.73	5.76	8.19
900x900	3.69	5.70	8.12
1000x1000	3.65	5.65	8.05
1100x1100	3.62	5.60	7.99
1200x1200	3.58	5.55	7.93
1300x1300	3.55	5.51	7.93
1400x1400	3.53	5.47	7.82
1500x1500	3.50	5.43	7.77
1600x1600	3.48	5.43	7.72
1700x1700	3.45	5.37	7.68
1800x1800	3.43	5.34	7.64
1900x1900	3.41	5.31	7.60
2000x2000	3.40	5.28	7.56

TABLE VI
PRACTICAL SPEED UP RATIO FOR INFORMATION RETREIVAL FROM FILES WITH DIFFERENT SIZES USING MATLAB

Files size	Speed up ratio (n=20)	Speed up ratio (n=25)	Speed up ratio (n=30)
100x100	7.88	10.75	14.69
200x200	6.21	9.19	13.17
300x300	5.54	8.43	12.21
400x400	4.78	7.45	11.41
500x500	4.68	7.13	10.79
600x600	4.46	6.97	10.28
700x700	4.34	6.83	9.81
800x800	4.27	6.68	9.60
900x900	4.31	6.79	9.72
1000x1000	4.19	6.59	9.46
1100x1100	4.24	6.66	9.62
1200x1200	4.20	6.62	9.57
1300x1300	4.17	6.57	9.53
1400x1400	4.13	6.53	9.49
1500x1500	4.10	6.49	9.45
1600x1600	4.07	6.45	9.41
1700x1700	4.03	6.41	9.37
1800x1800	4.00	6.38	9.32
1900x1900	3.97	6.35	9.28
2000x2000	3.94	6.31	9.25