Scanning Process in Digitization of Records and Archives Materials

Nur Nazeera Asma Ruhaimi, Saiful Farik Mat Yatin, Nur Hajar Mohd Fadzil

Full Terms & Conditions of access and use can be found at
http://hrmars.com/index.php/pages/detail/publication-ethics

Scanning Process in Digitization of Records and Archives Materials

Nur Nazeera Asma Ruhaimi, Saiful Farik Mat Yatin, Nur Hajar Mohd Fadzil

Faculty of Information Management, Puncak Perdana Campus, Universiti Teknologi MARA, (UiTM), UiTM Selangor, Malaysia

Abstract
This academic writing is all about the scanning which is the types of scanning and types of scanners. There are several types of scanner that have been stated in this academic writing which is flatbed Scanners, Sheet-Fed Scanners, Integrated Scanners , Drum Scanners and Portable Scanners. This academic writing also have been discuss about the hardware and software that been used in the scanning process. There are also several issues or problems about the scanning that have been stated in this paper. There are also a few example of scanning resolution, In addition, there are also explanation about the scanning process and new technology in the scanning.
Keywords: *Scanning, Scanner, Scanning Process, Issues of Scanning, Problems of Scanning.*

Introduction of Scanning
Scanning is the first step to putting documents in electronic format. Through scanning can retrieve printed and digitized documents to work with texts and images to prepare documents for presentations and the public on the internet. For image terms that have different items the first is the result of the scanned document for example, the item on the scanner is the printed document and the picture you created on the screen you have checked the document is a picture. The second item is almost nothing images in non-text documents. Which of these terms includes drawings, pictures, image maps, and charts (Anderson and Maxwell, 2004).

Types of Scanning
There are three (3) types of scanning which is, text, photo, slides and negatives. Firstly, for **text** was mentioned by Simske and Baggs (2004). Text is historically the most important information in a document, that were allowed OCR and ICR (handwriting) technologies to provide repurposing, summarization, categorization ad indexing of the captured material. That is very helpful to using because OCR can save the more time and the money. It very highly recommendation to used it, otherwise OCR is good because it make fewer mistakes and was allows to save the time when editing the text files. In addition, by using OCR is extremely limited budget. OCR very helpful to user when

they want to used it. Others, text program is work with OCR software to create a text files and text programs are sometimes referred as work processors. Anderson and Maxwell (2004). Secondly is **photo** is part of the type of scanning. That including photos are a specialized region type, and they are of increasing importance as scanner purchasers focuses more on photography. It is very focused on photos as it is also one of the important types used for the process of scanning and even these items are also on some types of materials such as maps, drawings, photographs and others. And we also need to identify the features of the image before doing the scanning process that is to identify the content that can be done from top to bottom and from the bottom. This method is to be faster and easier to use. Last but not least, **slides and negative.** That is, provide a simpler analysis problem than photos; since they are surrounded by at least some width of black boundary on two or more sides sing this information, the process of finding these two regions is greatly sped up by first defining and conveying the black (sub-threshold) regions, and then forming regions from the non-black pixels within these regions.

Types of Scanner
Scanners are devices used to produce exact digital image replicas from photos, texts written on paper without altering them. This digitalization process can be stored as a file into the computer for reuse. In addition, the most commonly used scanners are scanners which store objects on the glass windows. This scanner has a variety of types available as an example flatbed scanner, sheet-fed scanners, integrated scanners, drum scanners and portable scanners. The scanned results will be obtained on the computer and images or texts will be found through the process

Flatbed Scanners

Sheet-Fed Scanners
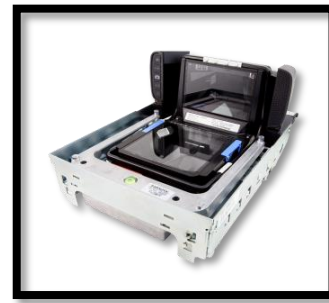
Integrated Scanners

Drum Scanners

Portable Scanners

Barcode scanner

Flatbed Scanners
First of all, a flat scanner is a commonly used scanner because it has functions used at home and office. How to use this scanner is a rolled mechanism under the document to get an image. And for those businesses they need high processing capabilities.

Sheet-Fed Scanners
This type of scanner is similar to that of a flat scanner, but the image is fed through the scanner. Others, this type are not used for books, but only one sheet that will use. This type of scanner most commonly used in homes and small offices.

Integrated Scanners
Integrated scanners are becoming one of the most modern types of scanners. It is comes to obtaining images. Which is ATMs feature this type of built-in scanner for check-processing and approval.

Drum Scanners
These types of scanners are used mainly for capturing a picture and producing at a very high-resolution rate. For instance, there are using to scan the maps that need using a good ability to do it. Not all companies that make these scanners, it just a few that only used it. Which is, see from the cost side, it takes the high cost of producing a scanner such as this. It is considered a tremendous upgrade to a regular flatbed scanner.

Portable Scanners
The portable scanner is designed to capture text and other data while on the go or lack of time. It makes user easier while using this scanner. This scanner is using the powered by batteries and the content result of the scan will be stored on the portable scanner after that can transfer the content to a computer. It can to transfer the content it only using a cable wireless connection.

Hardware of Scanning
Hardware of scanning is a device that will be used for the scanning process to success. The term hardware will be describes the physical aspects of computers and related devices. Hardware will be using in different ways if the types of scanner is different such as for barcode scanner and for the photos scanner. Hardware scanner is an easy and simple way for the people to get the information and data from the windows PCs and servers connected to the network. People just to enter the IP subnet range that they wants to scan and hardware scanner bring back the host name, CPU speed, Hard disk size and serial number and the amount of physical memory in the systems. Then it neatly displayed and exportable directly to excel or CSV file. There is input hardware and output hardware for the scanner which is input hardware is a devices that translate data into a form the computer can process such as translate the words, number, sound and pictures into binary that of or on electrical signals or light pulses. Output hardware is device that translate information processed by the computer into a form humans can understand. It is translate binary code into words, number, sound and pictures.
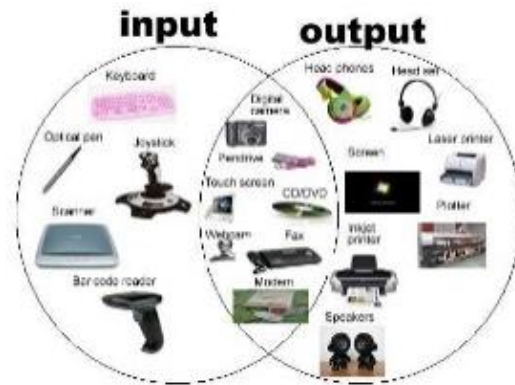
Figure 1: Scanning hardware

Figure 2: Input and output hardware for scanner

SOFTWARE OF SCANNING

Software is a general term for the various kinds of programs used to operate computers and related devices. There are many kind features of software for the scanning process. Software also often will be divided into application software which is programs that do a work users a directly interested in and system software are includes the operating and any program that supports application software. In the scanning there are many different types of software which is software for the photo editing, video, text editing, and others. To ensure that the scanning process be proper operation people have to install the driver and software. The driver and software are work together to provide seamless and optimal operation of the scanner. Some of the software that accompanies scanner have a limited in its capabilities and functions or it is trimmed down the version of a more powerful program and its required additional third-party software.

The scanner driver
The scanner driver is a needed to bridge the gap. It is also operate the scanner and transfer the digitize file to the hard drive or software. The scanner driver is a specialize version of the driver that is accessible through Photoshop, word or the others programs. The scanner driver also will be standalone and plug in. the standalone means the driver runs the scanner without involve with the other software and save the files to the hard drive. However, the plugin be opened within Photoshop or Word which is by choosing the import to insert the picture from the scanner and after the scanning process hand the files directly to the Photoshop or Word for immediate use. The scanner drivers have to group which is native and third party. Native Scan drivers are as a sketchy or robust as the manufacturer deems necessary and the features also can be vary from model to model. The third-party software means it is focus entirely on creating scan drivers which are oftentimes more robust and offer a better control over the scanner and also scanned the images than the native drivers
(Wentzel, 2006)

OCR software
OCR stand for Optical Character Recognition that people will use when they want to convert the archived images into the text with the word processors. The OCR is the better when the people want to save their time and money. The OCR program also makes a few mistakes and allows the users to save time when editing the text files. The two widely used and highly recognized OCR software

program are Omni Page Pro and also Prime Recognition. These two programs are accurate and robust than the WOCAR (Anderson and Maxwell, 2004)

FTP software
FTP or File Transfer Protocol program is a program that people use when they are ready to upload the files and project into the internet for the viewing. The FTP is a connected with the server and it allows users to create directories and move files into the appropriate folder which is where they can view on the internet. WS_FTP LE is excellent free FTP software that available for the institutions, students, teachers and also government and non-profit organization. (Anderson and Maxwell, 2004)

PDF software
PDF or Portable Documents Format is a widely used format that allows people to exchange the information instance where the people might not have the proper program to open the files or when the users does not want others to be able to modify and changes the files. PDF provides quick and easy solution for the getting files online viewing and downloading. The PDF should be used as a companion with the current formats of for enhance site accessible. For the PDF the Reader Acrobat is a leading software the creating PDF. It is because Reader Acrobat provides many options when create the PDF and its works with Microsoft Office applications which just click the button. Anyone with the computers can view by download if using the PDF software (Anderson and Maxwell, 2004)

InDesign
InDesign is software for the desktop publishing that have been created by the Adobe system. InDesign can be used to create a poster, flyers; brochures, magazines, newspaper and others InDesign also can publish content that suitable for tablets devices in conjunction with Adobe Digital Publishing Suite. However. InDesign also supports program such as EPUB, SWF and XML for use in other digital and online formats.

WOCAR
WOCAR will converts scanned documents to text documents, with support for OCR and the TIF, TXT and RTF formats, integrating several conversion options. WOCAR is an application that designed to help users convert scanned file to the text documents and it's also is capable of processing text files written in the English and French languages. WOCAR is a good program to use but the users have to consider the other options if the resources allow it is because the WOCAR have limited functionality and use. WOCAR be automatically default to the format so the users never needed the change any file setting (Opris, 2001)

Adobe Photoshop
Adobe Photoshop is a raster graphics editor that developed and published by Adobe Systems for macOS and Windows. Adobe Photoshop can edit and compose raster image in the multiples layers and support the mask, alpha compositing and the others several colour models including RGB, CMYK, CEILAB, Spot colour and duotone. The raster graphics has limited abilities to edit or render text, especially troughs clipping part. (Wikipedia, 2017a)

Issues of Scanning

The main issues that have faced by these technologies is space which means that the archive need space to store the large amount of data. Indeed there is also demand of efficient data compression technique that to provide uninterrupted services to the users. Issues of scanning also have been associated about the health and safety issues. It is because of the Acrylonitrile butadiene styrene (ABS) plastic which is a common thermoplastic polymer. It is glass transition temperature is approximately 105 °C (221 °F). ABS is amorphous and therefore has no true melting point. Due to the toxicity levels of ABS plastics it can produces fumes and the unpleasant smell and it can cause of the eyes and lung irritations. Issues for the staff is they does not have enough the strong technical staff for the scanning process and they need to build the strong technical staff to handle a new scanner technologies they also have to hired the experience staffs to handle the scanning process and the staff that have skill in the scanning process. (Groenendyk and Gallant, 2012)

Problems

First problems is, when equipment troubleshooting and logistics which is when hardware is essentially plug and plug where the 3d of the scanner is connected to the computers through the USB and then the software and associated drivers are installed through either with an online download or the supplied the CD-ROM. However the MakerBot Replicator will involve a much lengthier and complicated troubleshooting process. Second is, Equipment maintenance which is the regular maintenance for the MakerBot Replicator is low. The time involved is roughly fifteen minutes, and the cleaning task such as cleaning the build platform with rubbing alcohol, greasing the rods on which the printing head moves, and changing and reloading the plastic filament that task only needs to be done after three months of constant printing.

Scanning Resolution

Even though the concept can be confusing, a resolution is just a measurement of how many pixels a scanner can sample in a given image. The resolution that can be measured by a grid and it must be measured in the sample per inch or SPI. Then, many times people refer to resolution with the term DPI, "dots per inch", or with PPI, "pixels per inch". SPI, DPI, and PPI all describe the same thing. It commonly makes people confusing about that.

Table 1: Example of resolution

Resolution	Data Size	Print Size						
		4 × 6" (100 × 150 mm)	5 × 7" (130 × 180 mm)	8 × 10" (203 × 254 mm)	A4 (210 × 297 mm)	10 × 12" (254 × 305 mm)	A3 (297 × 420 mm)	13 × 19" (329 × 483 mm)
1200 dpi	25 MB	672 dpi	517 dpi	331 dpi	320 dpi	265 dpi	226 dpi	204 dpi
1600 dpi	44 MB	896 dpi	689 dpi	441 dpi	427 dpi	353 dpi	302 dpi	272 dpi
2400 dpi	99 MB	1344 dpi	1034 dpi	662 dpi	640 dpi	529 dpi	453 dpi	409 dpi
3200 dpi	222 MB	2016 dpi	1551 dpi	993 dpi	960 dpi	794 dpi	679 dpi	613 dpi
4800 dpi	395 MB	2688 dpi	2068 dpi	1324 dpi	1280 dpi	1058 dpi	905 dpi	817 dpi

The Scanning Process
Process of scanning is involved by two item firstly is **specialized software** which is a specific software that is the scanner driver needed to bridge the gap. Scan drives handle scanners and transfers file digitization to hard disk and software. Scan drivers may be standalone drivers or plug-ins, and it is also a specific version of the drivers that can be accessed through the Photoshop program, Word and others. In addition, scanning of the process is several steps that must be followed to achieve the desired results when implementing a digitization process. The process was captured started from hard-copy original received, which is it received the item then go through the make digitization or records determinations, after that it need to verify quality legibility & non-duplication' establish custody . Next is any special concern which is 2-side original handling, attachment and others. Go to the next process is establish file output format & apply preliminary metadata(per standards), so it proceed to document preparation which is remove stapler & binders, smooth folds & wrinkles, refresh damaged pages, apply barcode and insert target/ divider sheets. So to the next is digitization or scanning with (OCR); Output to Staging Drive and verification by pages, special, concerns, addressed & other, however it proceed to hard-copy original returned to custody procedures and release capture data from staging drive to production system and lastly is images files, OCR, metadata are available for searches, retrieval, sharing and management.

New Technology in Scanning
The new technology in scanning is 3D printing and scanning technology can also provide enhanced data visualization techniques, opening new methods for researchers to present data to audiences.3D scanning is more advance and very easy to use. 3D Scanners is a device that analyses objects of the world or the environment to collect data on its shape and may be an example of its appearance is colour. The collected data can then be used to construct digital three-dimensional models. New technology is sometimes that are new to scanner that more advance and give a lot of advantages. It also relate with 3D bar Code which is a 3D barcode actually a one-dimensional barcode that arises to the surface, but unlike the 2D sticker code. 3D barcodes are read by using the differences in height between the bars and spaces.

New technology in scanning is very good because it according to time passage by the item. 3D is more effective and will come out with good result. 3D barcode scanners can be used during the manufacturing process or in hand held form. The scanners are necessary in order to track the efficiency of the production process. 3D barcodes are being used in inventory and for purchases. In time 3D barcodes will drastically boost the savings for manufacturers.

New technology in scanning is give a lot of benefit which is it more advance to use and the usage of the technology may help is more fast and accurate. For example, 3D barcode scanner is use in shopping mall or manufacturing process to scan the item or the material faster and quickly. It not use manual process, and the 3D barcode is very efficient to become go in market for future. *It was mentioned by website qrscanner.us*, Barcode 2D has been very useful in some industries including retail. Sometimes barcodes are printed directly on products such as books and other times printed on the sticker placed on the product and can be removed after purchase or packing of the product. Others for 3D barcodes originally designed for use in the automotive industry to help detect vehicle parts during the manufacturing process. In addition, the 3D bar code has proven to be more effective as it can accommodate extreme chemicals and heat that are not available in retail environments.

Scan Quality
The typically colour of scanners is read RGB data from the array. The RGB means is red, green and blue. This data from array then processed with some proprietary algorithm to correct for different exposure conditions, and that will be sent to the computers via the device's input and output interface. The scanning array characteristics will influence the varies of Colour Depth. The colour depth that usually have been use at least 24 bits. And for the high quality models of the scanners will have 36-48 bits of colour depth. Resolution is another qualifying parameter for a scanner. And they used measurement in Pixels PerInch which is PPI or, sometimes the measurement more accurately referred to as Samples PerInch which is SPI. Interpolated resolution have been meaningful parameters and manufacturers like to refer to that parameters. Interpolated resolution is much higher because of software interpolation. The true resolution of scanners that have been determined by using resolution test charts is the one of the effective of the resolution. Flatbed scanners is a considerable as a lower than the manufactures given optical resolution by most all consumers and it's have been call as an effective resolution. The scanner noise, optical flare, poor analog to digital conversion, scratches, dust, newton's rings, out of focus sensors, improper scanner operation and poor software are can make the purity diminished. High end film scanners that use the large Kodak Tri-Linear sensors have to been followed if to produces the purest digital representations of the films said Drum scanners. The others important parameters for a scanner is Density Range which is Dynamic Range or have been call as a Drange. It is a high range which is that the scanners is able to record the shadow details and brightness details in one scan process. (Wikipedia, 2017a)

Computer Connection
The scan must be removed from the scanner to the computer or information and records storage system for further processing or storage. There are two basic issues in the scanning process which is first how the physically the scanners are connected to the computer and second how the application will gets the information from the scanners.

Direct physical connection to a computer

The scan file size can reach approximately 100 megabytes and the scanner can also generate volume data in seconds, and make the desired fast connection. Scanners can communicate with their host computer using one of the following physical interfaces. This lists about physical interface from slow to fast. First is the parallel port that connects through the parallel port is the slowest usual transfer method. This is the previous scanner that has a parallel port connection that cannot transfer data faster than 70 kilobytes per second. Second is GPIB which is a General Destination Bus Interface. This interface is only used by some scanners and largely serving documents or windows environment. The third is the SCSI Small Computer System Interface. This interface is rarely used since the beginning of the 21st century because the interface only supports computers either on cards or built-in. Next is Universal Serial Bus which is USB. This USB scanner can transfer data quickly. The fifth is Fire Wire or it has been called IEEE in 1394. The interface has a comparable speed of USB 2.0 but the device may not support all speed. The last in direct physical connection to the computer is proprietary. This proprietary interface has been used on some previous scanners who have used proprietary interface cards and not standard interfaces. (Wikipedia, 2017a)

Indirect (network) connection to a computer
In the 1990s professional flatbed scanners were available on local computer networks. This proved useful to publishers, print shops, and others. This function is largely used as a flatbed scanner cost which is reduced enough to make unnecessary partnerships. From the year 2000, versatility has become available for both office and small users, with the ability to print, scan, copy and fax in one single tool that can be provided for all working group members. Battery-powered mobile scanners store scans on internal memory; they can then be transferred to a computer either via a direct connection, usually USB, or in some cases a memory card can be removed from the scanner and installed to the computer (Wikipedia, 2017b).

Conclusion
Scanning technique has an important role in image processing and it actually affects the performance of the compression algorithm. A survey has been collected on different scanning approaches. We conclude that various scanning techniques are used to convert 2-D signals into 1-D signals and produce better compression ratios. As with computers, the internet, and important information technology that comes before them, 3D printers and scanners have the potential to significantly change the services of the library. Not only do they allow physical items to be stored in digital format, they can also physically create these items at the request of library visitors. Scanning really helps all users for this because it has various features and its use is very friendly to use.

References

Anderson, C. G., & Maxwell, D. C. (2004). *Starting a Digitization Center*. USA: Chandos Publishing.

Acrylonitrile butadiene styrene. (2017, December 17). Retrieved December 17, 2017, from https://en.wikipedia.org/wiki/Acrylonitrile_butadiene_styrene

Adobe InDesign. (2017, December 13). Retrieved December 17, 2017, from https://en.wikipedia.org/wiki/Adobe_InDesign

Adobe Photoshop. (2017, December 15). Retrieved December 17, 2017, from https://en.wikipedia.org/wiki/Adobe_Photoshop

Groenendyk, M., & Gallant, R. (2013). 3D printing and scanning at the Dalhousie University Libraries: a pilot project. *Library Hi Tech*, 31(1), 34-41. doi:10.1108/07378831311303912

Nicholson, S. W., Peiffer, R., & Shaw, J. D. (2011). Hardware in libraries: making informed choices. *Library Hi Tech*, 29(1), 73-82. doi:10.1108/07378831111116912

Simske, S. J., & Baggs, S. C. (2004). *Digital capture for automated scanner workflows*. Proceedings of the 2004 ACM symposium on Document engineering - *DocEng 04*. doi:10.1145/1030397.1030431

Wentzel, L. (2006). Scanning for Digitization Projects. *Library Hi Tech News*, 23(4), 11-13. doi:10.1108/07419050610674712

Wocar D. (2007, June 26). Retrieved December 17, 2017, from http://www.softpedia.com/get/Office-tools/Other-Office-Tools/Wocar.shtml