Document Imaging Systems

Nabil M. Hewahi, Mohammed Hannoush, Shady Mohana
Department of Computer Science
Islamic University of Gaza, Palestine
nhewahi@iugaza.edu

ABSTRACT

In this paper, we present concepts of document imaging systems and also a survey on some of well known products in this field. To make easy understanding for the studied products, we show a comparative study based on the basics of imaging systems as well as strategies used by those products. This study is very useful to help organizations to choose appropriate document imaging systems tools based on their needs.

Keywords: document capture, document management.

1 INTRODUCTION

We all need paper to do our work, but it accumulates quickly. Our files grow fatter and fatter, and then they grow some more. Folders and filing systems make it easier to find our documents. Records managers organize archive and retrieve our information. But the amount of paper keeps growing. Paper files are often hard to find. Records may not be in their proper folder. Or they may be “borrowed” and then lost on somebody’s desk. Studies show that professionals often lose up to 500 hours a year just looking for documents. Those days are gone. Document imaging offers a better way to manage the records you rely on. Document imaging systems should [7][12][19]:

1. Enable you to manage millions of records and retrieve the one you need in seconds.
2. Be a pleasure to use, whether you’re the person that needs the files or the records.
3. You share documents with colleagues while protecting confidential information.
4. Allow you to e-mail or fax files with the click of a mouse.
5. Provide an easy way to share documents with other offices or take them on the road.
6. Conform to the way you work, rather than forcing you to change.

2. WHAT IS DOCUMENT IMAGING?

Document imaging is the conversion of paper documents into electronic images on your computer. Once on your desktop, these documents can be retrieved effortlessly in seconds. Every organization generates large amounts of paper and electronic documents. We have all developed our own ways to store important files, yet things continue to get misplaced. Everyone knows the frustration of not being able to find a file right when we need it most. Traditional methods of storing paper and electronic records require a great deal of effort to manage, distribute and find those documents. As the number of files grows, the time and effort required to manage them also increases. Document imaging revolutionizes the archival of information and provides the means to rapidly find retrieve and share all documents in your system. Just as the Internet has boomed in popularity because of the fast access it provides to information stored in web pages, document imaging systems provide tremendous value because of the fast access they provide to information stored within an organization’s documents.

Document imaging builds on the strong points of paper documents: Files are scanned or electronically converted and a high-resolution photocopy is stored on a hard drive or optical disk. Electronic “index cards” can attach information to a document such as author, reference number or date created. Files can still be viewed, printed, shared and stored, but imaging adds an enormous advantage by giving documents active content.

3. BENEFITS OF DOCUMENT IMAGING

Thousands of organizations around the world use document imaging every day instead of paper filing systems. Document imaging offers a number of benefits over paper or microfilm systems [7][19].

a. Fast retrieval: Imaging lets you find documents quickly without leaving your desk. Paper and microfiche are slower because users must go to files and search manually.

b. Flexible indexing: Imaging can index documents
in several different ways simultaneously. Indexing paper and microfilm in more than one way is awkward, costly and time-consuming.

c. Full-text search: Imaging systems can retrieve files by any word or phrase in the document, a capability that is impossible with paper or microfiche.

d. No lost files: Imaged documents remain in their folders when being viewed, so none are lost or misplaced. Plus, index template and full-text searches can find documents if they are accidentally moved. Lost documents are expensive and time-consuming to replace.

e. Digital Archiving: The risk of loss or damage to paper or electronic records is reduced with a document imaging system. Keeping archival versions of documents in a document imaging system helps protect paper documents from over-handling and keeps electronic documents in a non-proprietary format.

f. Share files easily: Imaging makes it easy to share documents electronically with colleagues and clients over a network, on CD or through the Web. Paper documents usually require photocopying to be shared, and microfilm requires conversion to paper.

g. Improved security: Imaging can provide better, more flexible control over sensitive Documents. Imaging controls security at the folder, document or individual word level for different groups and individuals. In contrast, all paper documents in a filing cabinet or filing room have the same level of security.

h. Save space: Imaging will help recover valuable office space that was previously taken up by bulky paper files.

i. Disaster recovery: Imaging provides an easy way to back-up documents for offsite storage and disaster recovery. Paper is a bulky and expensive way to back-up records and is vulnerable to fire, flood and theft.

No longer just ink on a page, document text is “read” by Optical Character Recognition (OCR) technology. A system should allow you to Retrieve files by searching for any word or phrase in the text, by folder location or by “Index card” information. Which documents people can read, and what actions or modifications they can perform on these documents, depends on their level of security, which should be controlled by the document imaging system.

Selecting the right document imaging system can be an exciting task. There are many aspects to consider ensuring that it fits your organization’s needs. Following is a description of the five basic components of what to look for when choosing your system.

4. THE BASICS OF DOCUMENT IMAGING

Any document imaging system has the following basic components [6] [7][12][19][20][22]

1. Scanning and importing tools to bring documents into the system.
3. Indexing systems to organize documents.
4. Retrieval tools to find documents.
5. Access control to provide documents to Authorized people.

4.1 Bringing in Documents

There are three primary methods of bringing files into a document imaging system

1. Scanning, for paper files.
2. Conversion, for creating unalterable images of electronic documents.
3. Importation, for creating modifiable versions of electronic documents.

4.1.1 Scanning

Scanning a document produces a raster (picture) image that can be stored on a computer. When choosing a scanner, it is important to consider overall budget and the size and volume of paper to be scanned. The ability to use a wide range of scanners is one of the defining characteristics of a good imaging system.

A document imaging scanner should have an Automatic Document Feeder (ADF). This device allows stacks of paper to be placed into a tray and automatically fed one page at a time into the scanner, speeding up the scanning process. Scanners can handle a variety of paper sizes to cope with all the requirements. Other options such as color or grayscale, the speed of document scanner and the availability of simplex and duplex modes increase the scanner’s price.

If there is a large volume of documents to scan, i.e. thousands or millions of pages, it may be more practical and economical to use an outside scanning service bureau. To support this option, the imaging system must accommodate easy database synchronization between information scanned by the service bureau and pages scanned in-house. The data volumes containing images and index information need to be modular and easily portable. This ensures
that the documents scanned by the service bureau can be incorporated into a “live” system without interrupting or re-indexing existing work. This option is often referred to as “portable volumes.” If an organization has several offices and needs to share the documents scanned by each, portable volumes capability provides an easy Way to distribute files.

4.1.2 Conversion

Converting documents is the process of transforming electronic word processor or spreadsheet documents into a permanent raster image format for storage within an imaging system. Windows applications, such as Microsoft Word, Excel or AutoDesk AutoCAD, can “print” existing files into an unalterable image of the document. These images are usually stored as archival-quality TIFF (Tagged Image File Format). The conversion process also generates a complete text file, while retaining the visual Formatting and layout of the original file. This text file can then be used for full-text indexing of the document to assist with later retrieval.

Converting electronic documents bypasses scanning, saves paper and printing resources, and produces a cleaner image than scanned paper files. This method of “imaging” electronic documents is best suited for permanent archives.

4.1.3 Importing

Importation, also known as electronic document management, is the second method for bringing electronic files, such as office suite documents, graphics, audio clips or video files, into a document imaging system. Files can be “dragged and dropped” into an imaging system but are modifiable and remain in their native format. These files can be viewed in their original format by either launching the originating application or by using an embedded file viewer from within the imaging system.

4.2 Storing Documents

Once brought into the system, documents must be stored. Document imaging storage systems must encompass changing technologies, increasing numbers of document volumes and the tests of time. The needs and budget for image storage are best determined by the individual organization involved.

A good document imaging system should be able to use any storage device currently available—as well as those on the horizon—to provide long term document storage. This allows you to select the equipment that best meets your needs, both now and in the future.

To ensure readability in the future, if a document imaging system is to be used for digital archiving, the files should be stored in a nonproprietary format. The computer industry advances so quickly that storing document images or text files in a proprietary format may leave a company held hostage to the fortunes and whims of a single company. At this time, there are five primary storage options: Magnetic Media (Hard Drives), Magneto-Optical Storage (MO), Compact Discs (CD), Digital Video/ Versatile Disk (DVDs) and Write Once Read Many (WORM).

4.3 Indexing Documents

When paper documents are received in an office, they must be organized to be useful. They are usually labeled, sorted, indexed, stapled, placed in folders and filed in a cabinet. Without these steps, nothing could be found in a busy workplace. Electronic documents are no different. A document imaging system should provide several different methods of organizing information for future use. Whatever combination of indexing methodologies is used, it needs to be easily used and understood by the people who retrieve the documents, as well as those who file them.

There are several schools of thought about how much change to filing methodologies should be introduced. In general, the more a document imaging system can adapt to existing procedures, the less upheaval and training is involved, and the greater the likelihood the system will be used on a regular basis.

There are three primary ways to organize documents in an imaging system: index fields, full-text indexing and folder/file structure. For more details, refer to [19]

4.4 Retrieving Documents

Retrieval is where a powerful indexing system pays off. Users need to be able to use common sense tools to find any document within the system based on what they know. In some cases, this means browsing through folders, in other cases it could mean conducting index field searches. If all that is known about a needed document is a word or phrase it contains, a full-text search would help find the relevant file. Whatever the method, document retrieval must be simple and user-friendly.

Users who are familiar with a document’s text should be able to use that information to find what they need. Some systems can only find pages based on indexed keywords. This method is not always
helpful because the person who selected the keywords may not be the one searching for the file. To be truly useful, a document imaging system must use full-text retrieval.

Similarly, using the document name and folder view to find a document can be helpful and intuitive, but is not always the best method. Sometimes a person will know exactly which document they need, but not know what folder it was placed in or how the document was named. Using index field information to find a particular document can also be helpful. A full featured imaging system will have user-definable template fields. Index field searches allow users to comb through millions of records in seconds to find their needed documents. Of course, a person will need to know how the document has been categorized and what index fields have been assigned to it.

To maximize search effectiveness, an intelligent search system should be able to combine template searches with full-text searches with document or folder name searches into one comprehensive search. A good imaging system makes retrieval of relevant documents fast, easy and efficient.

4.5 Controlling Access

The final mandatory component of a document imaging system is the access control. In many computer environments, different people use different types of computer equipment from different locations to search for information. A full-featured imaging system must provide these different users with appropriate levels of access, without compromising confidentiality or security. To do this, a system must have two fundamental features, broad availability and comprehensive security[19].

4.5.1 Broad Availability

An imaging system must offer several different ways of accessing files. A broad level of access saves limited financial resources, intellectual capital and network bandwidth. The most common method of access is through the user's desktop. Every document imaging system must provide a client-based user interface that enables the scanning, indexing and retrieval of documents. Without this basic interface, the System cannot function. To provide broad availability and access flexibility, imaging systems must meet the requirements of offices with diverse uses and remote locations. Document imaging is no longer an "in-the-office" process. Many users require portability to exchange documents with other colleagues, or to work off site. This is frequently done through CDs, notebook computers, or e-mailing of documents. Imaging systems without this flexibility limit the abilities of the user. In addition, sharing documents through the Internet or an intranet allows system administrators to deploy an imaging system across their entire network or even to the public. Users should be able to search, retrieve and view documents with any web browser. Browser-based document access removes limitations of location and computer platform (Windows, Macintosh, UNIX, etc.).

4.5.1 Comprehensive Security

As organizations use imaging systems to archive a larger variety of documents, both public and private, a system of access control needs to be present. A comprehensive security system must allow the system administrator to control what folders and documents users can see, and what actions they can perform on those documents (Edit, copy, delete, etc.). This system must control access to folders, documents and even redacted images and text in a simple and complete manner. The ability to deploy imaging to a wide variety of users requires a robust security system combined with an elegant user interface. A good access system will make document imaging available to every authorized person, whether in an office, at a remote location or over the Web, all without compromising system security.

5. IMPLEMENTING DOCUMENT IMAGING

This includes several steps and should be performed in the following sequence
2. Scaling from pilot project to enterprise solution.
3. Installations.
4. Training, this includes end user, system administration, implementation consulting and supervised hand-on operation.

6. DEFINITIONS AND TERMS

In this section, we present some definitions and terminology concerning the document imaging systems.

Document Capture : Document capture is the process of converting paper documents into digital images and index data. Images are stored as data files on a storage system, and the indexes are stored in a content management system. In a document capture environment, the primary deliverable to the back end content management system is the image. The index information is provided so users can retrieve the original image. Without the image, the index information has no value[1] [4] [6] [16].
Data capture: (historically referred to as forms processing) is the process of automatically extracting information from forms. In a data capture environment, the information contained on the form is the primary deliverable to a back end line of business application or a legacy database. Oftentimes the images themselves are discarded once the desired information has been extracted. Most forms oriented products available on the market today require the user to redesign their forms in order to achieve acceptable recognition rates. However, new technologies being released into the market today can work effectively with existing real-world forms while maintaining adequate recognition accuracy [1] [4] [6] [16].

Production Capture (Merging Document and Data Capture): Merging document capture and data capture into a single product is a natural evolution of production capture. The result is a product offering that is more intuitive to use and easier to deploy, especially for users who find themselves having to deal with form and document requirements within the same department or enterprise. By merging the two product segments into a single product offering, it is now possible to deploy a single application to meet a wide range of capture applications that in the past would have required the deployment of two or more applications [16].

Capture Flow: Is the process which allows you to effectively design an enterprise-specific information capture solution, once that accepts your existing technologies and adapts to current and future needs [4].

7. BACKGROUND

In the next sections, we shall make a comparative study among three well known document imaging systems. Those three have been chosen from several document imaging systems in the market. The chosen products are LaserFiche[17][18], Accent Capture[15] and InputAccel[4]. Several well known companies produced products for document management and imaging systems. Haltsystems produced three document imaging systems, each is specialized in a certain field, eMedRec for medicine, items flow for financial institutions and OPTICA for business solutions [11]. Tokairo produced a software tokairo to provide solutions for document management and education systems[23], and Image[1] and DIS[2] for general document imaging applications. Other general purpose imaging systems are as IMR Alchemy from Alchemy [21], PMAS from elibrary[8], FINEOS from FileNet[9], liberty IMS from liberty IMS[20]. Some other products are produced from legato [13], OnBase[21], Documentum [22], optimissystems [6], and cortech[4].

8. COMPARATIVE STUDY

In this section, we present a comparative study to differentiate among the three chosen products. The study includes various points to be the basics for the comparison among several points in the document imaging where to observe the differences among the products. We can distinguish among the products based on the basics of document imaging systems and the product strategy. The basics of document imaging are based on bringing, storing, indexing retrieving and controlling access. The product strategy may include system details, the strategy, document options, integration, distribution, workflow, security and scalability.

8.1 Basics of Document Imaging

8.1.1 Bringing

Table 1 shows the bringing differences of the three products based on bringing.

8.1.2 Storing

Table 2 shows the difference of the three products based on storing.

8.1.3 Indexing

Table 3 shows the differences of the tree products based on indexing.

8.1.4 Retrieving

Table 4 shows the differences of the three products based on retrieving.

8.1.5 Controlling Access

Table 5 shows the differences of the three products based on controlling access.

8.2 Product Strategy

8.2.1 System Detail

Table 6 shows the differences of the three products based on system details.

8.2.2 Document Options

Table 7 shows the differences of the three products based on document options

8.2.3 Distribution

Maximizing the value of information locked in
an organization’s documents requires putting it in the hands of decision-makers when and where it’s needed most. From staff in remote offices, to clients, the public and regulatory officers, individuals within and beyond your organization rely upon fast information access to work at peak productivity. Table 8 shows the differences of the three products based on distribution.

Table 1. the differences of the three products based on bringing.

<table>
<thead>
<tr>
<th>Product / property</th>
<th>Scanning</th>
<th>Importing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LaserFiche</strong></td>
<td>Scan paper with Kofax supported scanners.</td>
<td>LaserFiche import agent, who knows where to look for documents, how to organize docs and how to minimize network traffic. LaserFiche snapshot imports files without the need of scanners. Manually and automatically imports documents for storage. The import list feature allows users to automate the process of importing documents into a LaserFiche database; this feature allows more flexibility than the standard import feature.</td>
</tr>
<tr>
<td></td>
<td>Auto name documents while scanning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scan multiple pages.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Automatically cleanup pages.</td>
<td></td>
</tr>
<tr>
<td><strong>Accent capture</strong></td>
<td>Scan paper with Kofax supported scanners.</td>
<td>Import images and data documents from any scanner.</td>
</tr>
<tr>
<td></td>
<td>Drives both simplex and duplex scanners.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Detection of errors during scanning.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACIS enables remote scanning of documents.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Automated recognition techniques such as OCR,ICR,OMR or bar code to read zones</td>
<td></td>
</tr>
<tr>
<td><strong>InputAccel</strong></td>
<td>Multistream image (color and black and white).</td>
<td>Import images and data from any scanner.</td>
</tr>
<tr>
<td></td>
<td>Point and click user interface.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Support more than 300 scanners.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indexing fields at all levels.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drives both simplex and duplex scanners.</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. the differences of the three products based on storing

<table>
<thead>
<tr>
<th>Product/ property</th>
<th>Storing</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaserFiche</td>
<td>CDs, HD, FD, DVDs, WORM</td>
</tr>
<tr>
<td>Accent capture</td>
<td>The same as LaserFiche</td>
</tr>
<tr>
<td>InputAccel</td>
<td>The same as LaserFiche</td>
</tr>
</tbody>
</table>

Table 3. the differences of the three products based on indexing

<table>
<thead>
<tr>
<th>Product / property</th>
<th>Index fields</th>
<th>Full text indexing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent Capture</td>
<td>Database can be defined for one or more index fields. Web Validation Server,</td>
<td>Barcode indexing can be used to separate documents by type if available.</td>
</tr>
<tr>
<td></td>
<td>allows docs to be indexed remotely by anyone with an internet connection.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use bar codes to automate indexing.</td>
<td></td>
</tr>
<tr>
<td>InputAccel</td>
<td>Indexing of page level can be achieved using IQA (Image Quality Assurance).</td>
<td>Barcode indexing can be used to separate documents by type if available.</td>
</tr>
<tr>
<td></td>
<td>IQA also helps users index unstructured documents such as invoices.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Up to 20 index zones and associated fields per page.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>InputAccel also provides DocID (Document ID) for automatic document classification.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auto indexing using barcode.</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. the differences of the three products based on retrieving

<table>
<thead>
<tr>
<th>Product / property</th>
<th>Retrieving</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaserFiche</td>
<td>LaserFiche server engine based on LaserFiche API commands that communicate with this server engine. LaserFiche COLD which is fully searchable reduces the time spent retrieving documents from physical storage spaces. The LaserFiche search screen provides a very intuitive way of searching and retrieving documents based on what the user knows, whether the user knows the name of the folder containing the document, the name of the document, the index information or even a unique phrase in the document. Fuzzy logic searches find documents with misspelled words or OCR errors. Search results can be saved in folders for quick referencing and easy access. Browser, name and text searches are all very simple and user-friendly methods of searching and retrieving information from the LaserFiche database. View the word in context of the sentence before retrieving the entire image.</td>
</tr>
<tr>
<td>Accent Capture</td>
<td>Ascent capture internet server.</td>
</tr>
<tr>
<td>InputAccel</td>
<td>1. InputAccel server.</td>
</tr>
</tbody>
</table>

Table 5. the differences of the three products based on controlling access

<table>
<thead>
<tr>
<th>Product / Property</th>
<th>Broad availability</th>
<th>Comprehensive security</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaserFiche</td>
<td>Not available.</td>
<td>Comprehensive security controls access by user and function. In LaserFiche WebLink there is no comprehensive security active X controls. In LaserFiche united, the comprehensive security protects the entire digital archive from unauthorized access.</td>
</tr>
<tr>
<td>Accent Capture</td>
<td>Not available.</td>
<td>Event Logging is maintained at both central site and remote site.</td>
</tr>
<tr>
<td>InputAccel</td>
<td>No available.</td>
<td>No information (or not available)</td>
</tr>
</tbody>
</table>

8.2.4. Integration

A complete set of integration tools that streamline custom application development and integration with existing technology. Designed with open architecture, the Integrator’s Toolkit provides the capability of affordably and quickly customizing software solutions to your business
needs. Developers familiar with COM calls will be able to take full advantage of the Integrator’s Toolkit. Any Windows-based development language, such as C++, Access Basic or Visual Basic can be used to create applications. Visual Basic knowledge is helpful for modifying the source code and for incorporating customized commands. With the comprehensive documentation included in the Toolkit, detailed instructions and sample source code, information technology personnel have maximum flexibility to meet the specific business needs of a wide variety of organizations. Table 9 shows the differences of the three products based on the integration.

8.2.3. Workflow

Any organization's knowledge workers search for information, act on it, move it and archive it every day. This process, with the manual searching, faxing, photocopying and hand distribution, is costly and time consuming. The inefficiencies of the process divert your staff from the crucial business of making productive use of the information. A successful workflow solution automates this paper shuffle, effectively routing information to appropriate persons while being flexible enough to handle exceptions to the rules. It saves time and money by reducing photocopying, hand assembly line, but as responsible knowledge workers whose time is better spent making delivery and repetitive dragging and dropping. It automatically reminds staff of required tasks and notifies supervisors of action and inaction. Table 10 shows the differences of the three products based on the workflow.

8.2.6. Security

Expectations of security:

There are three proverbs of network security
1. Badly administered security regimes can be as detrimental as no security at all.
2. No human mind can create a puzzle that another human mind cannot solve.
3. The only absolutely secure network is one inch of space.

Table 11 shows the differences of the three products based on the security.

8.2.7. Scalability

The scalability of a system determines how much the imaging system can grow with your organization’s needs. For full scalability, a system should have the following attributes:
1. Support an entire enterprise’s users concurrently
2. Store all documents in the enterprise
3. Robust system architecture
4. Store information across multiple drives or servers
5. Support multiple databases
6. Expand to the Web
7. Publish information to CD or DVD

Table 12 shows the differences of the products based on scalability.

Table 6. the differences of the three products based on system details.

<table>
<thead>
<tr>
<th>Product / Property</th>
<th>System detail</th>
</tr>
</thead>
</table>
| LaserFiche         | LaserFiche server.  
|                    | LaserFiche Client access.  
|                    | LaserFiche workflow suite.  |
| Accent capture     | Central site.  
|                    | Remote site.  |
| InputAccel         | InputAccel server.  |
Table 7. the differences of the three products based on document options

<table>
<thead>
<tr>
<th>Product / Property</th>
<th>Document Option</th>
</tr>
</thead>
</table>
| LaserFiche         | View image and text side-by-side  
|                    | Zoom image to 100x magnification  
|                    | Anti-aliasing/grayscale for clearer display of text and graphics  
|                    | Print or fax images and text files  
|                    | Edit document text or export to your favorite word processor  
|                    | text files and TIFF Group III,TIFF Group IV,PCX and PDA  
|                    | GIF,BMP,JPEG,CALS images can also be imported into LaserFiche  
|                    | OCR can be performed on demand or as part of a batch  
|                    | Export text or images at any time  
| Accent Capture     | Full text OCR.  
|                    | Robust batch management.  
|                    | Automated extraction for OCR,ICR, or OMR.  
| InputAccel         | TIFF,PCX, bitonal images, color and grayscale,JPEG AND PDF image formats can be imported into a batch.  
|                    | Reading of patch code or bar code.  

9. CONCLUSION

In this paper, we presented an overview of document imaging concepts. We have also chosen three software products that can solve the problems of the document imaging (LaserFiche, Accent Capture and InputAccel) for a comparative study. A comparative study has been presented to distinguish among the three chosen products. This study helps any organization to choose one of those products based on its demands. The comparative study has been done on the basis of two main concepts, the basics of document imaging and the system strategy. One of the main problems in the available document imaging systems is that most of them do not support multi languages such as Arabic and Chinese and those which support lack a lot of capabilities. Some of the future direction for this research could be: 1. Widen the chosen products to be included it in the comparative study. 2. Considering document imaging software supporting Arabic / Chinese languages in a separate study and find the points (capabilities) where they should be developed to be put together to develop one reliable software. 3. If our criteria for imaging products proves to be a good scale, it would be easier to use it to be the scale for other imaging systems comparisons.
Table 8. the differences of the three products based on distribution.

<table>
<thead>
<tr>
<th>Product / Property</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LaserFiche</strong></td>
<td>LaserFiche webLink that provides a secure document access to remote offices, via standard web browsers. LaserFiche plus, CD publishing offers document portability. LaserFiche Email plug-in allows instant electronic document distribution via standard email application. LaserFiche COLD/COLD Plus, archives machine generated reports to digital media, eliminating great quantities of paper reports. LaserFiche COLDlink, delivers maximizes COLD productivity benefits with single interface access to COLD and LaserFiche imaging databases.</td>
</tr>
<tr>
<td><strong>Accent Capture</strong></td>
<td>Accent capture internet server (ACIS) for internet or intranet configuration. ACIS installed on the central site and acts as a wall between the central accent capture database and remote locations. Centralized administration which can manage the configuration of information, and also schedules for uploading batches. The central site's web server can be configured easily using Microsoft information server (IIS). Remote stations can be configured using Microsoft internet explorer. Authentication and encryption settings can be configured using IIS, the secure socket layer SSL can be enabled for encrypted transfer for data between the central site and remote site. Offline capability, with Accent Capture, remote stations do not have to be connected to the central Accent Capture server. Customizable web pages, the interface of the software that runs from remote locations is in the form of ASP and HTML pages, so you can use any HTML editor or even notepad to change the look of the pages to suite your own needs.</td>
</tr>
<tr>
<td><strong>InputAccel</strong></td>
<td>InputAccel server. InputAccel server can acts as a virtual manager and automatically routes data to the appropriate clients. InputAccel server is capable of automatic workload balancing, which will ensure you are getting the highest level of productivity. InputAccel server allows you to create a complex and a diverse capture flows, and maintained the integrity of data being routed through each. InputAccel server runs under the windows platform (NT, 2000, XP), and utilizes the power of windows to improve performance. InputAccel allows the incorporation of third-party modules into any and all capture flows, resulting in a fully customized, enterprise specific information capture solutions.</td>
</tr>
</tbody>
</table>
**Table 9.** the differences of the three products based on integration

<table>
<thead>
<tr>
<th>Product / Property</th>
<th>Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaserFiche</td>
<td>The open architecture and published database structure of LaserFiche, combined with the tools provides in the integrator's toolkits. The LaserFiche integrator's toolkit provides the capability of quickly and cost effectively developing custom imaging applications tailored to your business needs. The integrator's toolkit provides all necessary software and documentation to integrate LaserFiche with the third-party applications. With comprehensive documentation included in the toolkit, detailed instructions and sample source code, information technology has maximum flexibility to meet the specific business needs of a wide variety of organizations. The web based API provided with web Link simplifies integrating LaserFiche with any web application and does not require any client side source code.</td>
</tr>
<tr>
<td>Accent Capture</td>
<td>Accent Capture integrates quickly and easily into most popular workflow and content management systems. Including applications from IBM, FileNet, documentum, PC DOCS, Optika, icomXpress, Open text, IMR, Excalibur, Eastman software and others.</td>
</tr>
<tr>
<td>InputAccel</td>
<td>InputAccel is API to API compatible with enterprise content management systems from vendors such as IBM, documentum, OpenText, FileNet and eiStream. All key features such as initializing a workflow or mapping a document attributes are supported, allowing you to take complete advantage of all the strengths of your existing systems. Technologies from third-party providers can be easily integrated into InputAccel. Some of the features that the InputAccel provides for the documentum are, graphical browsing for the documentum database, mapping document-level index values to documentum attributes, and also the ability to attach a documentum workflow.</td>
</tr>
</tbody>
</table>
Table 10. the differences of the three products based on the workflow.

<table>
<thead>
<tr>
<th>Product / Property</th>
<th>Workflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaserFiche</td>
<td>LaserFiche Workflow suite, which increases productivity by automating document centered work process. The workflow suite include the following: server automation, work distribution manager, SMTP notification and finally audit trail. The Workflow suite provides the efficiency and security of rules-based routing and monitoring while also support ad hoc participation in the workflow environment. The workflow suite provides the stability and ease of use shared by all LaserFiche systems. The workflow suite's services copy and move documents via your computer network so you don't have to. The advanced audit trail option monitors and provides custom reports on user activity.</td>
</tr>
<tr>
<td>Accent Capture</td>
<td>Administrator controls how much of the capture workflow occurs locally, and how much elsewhere on the network. In some cases, it may make sense for the local subject experts to index and validate documents while in others, the origination site may scan.</td>
</tr>
<tr>
<td>InputAccel</td>
<td>InputAccel allow you to create customized CaptureFlows for your incoming data. InputAccel server allows you to create complex and diverse CaptureFlows, and diligently maintains the integrity of the data being routed through each. The InputAccel server acts as a virtual manager and automatically routes data to the appropriate clients, the clients perform the work while the server oversees them. Only InputAccel allows for the incorporation of third-party modules onto any and all CaptureFlows resulting in a fully customized.</td>
</tr>
</tbody>
</table>

Table 11. the differences of the three products based on security

<table>
<thead>
<tr>
<th>Product / Property</th>
<th>Security</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaserFiche</td>
<td>Comprehensive security controls access by user and function. LaserFiche webLink allows you to take advantage of the security setup on your LaserFiche database. The security on the LaserFiche database combined with the flexibility of LaserFiche WebLink security, allows the administrator to determine how users will log into the LaserFiche database.</td>
</tr>
<tr>
<td>Accent Capture</td>
<td>In Accent capture internet server, authentication and encryption settings can be configured using IIS, the secure socket layer protocol can be enabled for encrypted transfer for data between the central and remote sites. Administrator controls user access down to document type, secure 128 bit data stream encryption (Accent capture internet server)</td>
</tr>
<tr>
<td>InputAccel</td>
<td>InputAccel server works under windows platform (Windows NT, 2000, and XP). It uses the power of windows to increase the performance of the security. InputAccel server utilizing windows domain security, so administrators need not to worry about enforcing security at each and every workstation.</td>
</tr>
</tbody>
</table>
Table 12. the differences of the three products based on scalability

<table>
<thead>
<tr>
<th>Product / Property</th>
<th>Scalability</th>
</tr>
</thead>
</table>
| LaserFiche         | LaserFiche is optimized and communicates with the Microsoft SQL server database platform, the majority of application logic is located within SQL server as stored procedures to allow for superior platform scalability.  
N-tier architecture, which is the method of distributed computing in which a processing of a specific application occurs over n number of machines across a network. The primary advantage of this architecture is the scalability.  
With LaserFiche open architecture, the modular design and scalability will meet the changing you need. |
| Accent Capture     | Small standalone systems can be quickly and easily expanded to meet the most demanding throughput requirements.  
Accent capture internet server permits unlimited incremental growth at any remote site and in overall solution capacity. Note that the largest distributed applications in production have up to one thousand remote sites and capture up to one million papers per day. |
| InputAccel         | InputAccel can be deployed at a departmental and be scaled up easily to an enterprise level. The system's modular architecture is designed to expand compartmentally.  
One existing new feature that likely will be in release 5.0 is clustered servers that provide the following benefits, high availability, and scalability.  
InputAccel's eight signature strengths offer you the scalability to fulfill your specific requirements                                      |

10. REFERENCES

[1] www.image.com (image software, inc.).  