

Automatic File Management Sorting Using Lossless Data Compression Algorithm

Anna Liza A. Ramos
Saint Michael's College
of Laguna
Old National Road, Brgy.
Platero, City of Biñan,
4024, Laguna
annakingramos@yahoo.c
om.ph

Karl Hendrix M. Zarate
Saint Michael's College
of Laguna
Old National Road, Brgy.
Platero, City of Biñan,
4024, Laguna
karl.hendrix.zarate@gapp
s.smcl.edu.ph

Patrick Neil A. Reyes
Saint Michael's College
of Laguna
Old National Road, Brgy.
Platero, City of Biñan,
4024, Laguna
patrick.neil.reyes@gapps.
smcl.edu.ph

Ivan Paul D. Crisol
Saint Michael's College
of Laguna
Old National Road, Brgy.
Platero, City of Biñan,
4024, Laguna
ivan.paul.crisol@gapps.s
mcl.edu.ph

Khian Nicole Sison
Saint Michael's College
of Laguna
Old National Road, Brgy.
Platero, City of Biñan,
4024, Laguna
khian.nicole.sison@gapp
s.smcl.edu.ph

John Ace A. Mesoga
Saint Michael's College
of Laguna
Old National Road, Brgy.
Platero, City of Biñan,
4024, Laguna
john.ace.mesoga@gapps.
smcl.edu.ph

Abstract: File Management is a reflection of disciplined of how people organize things however not all people have this kind of attitude and even they want to make it organize they fail to it because of some interruption, time consuming to prepare and think where to save and what folder name to create and another things is most of the users file naming is not identifiable or not specifically name based on its contents. In this study, the researchers would like to enhance the existing built windows file management system capability through automatic file sorting management by year of creation, file type classification, folder naming category based on its file name content, elimination of file duplication, integration of file compression algorithm and file protection security. To realize the study, the researchers conducted a systematic investigation using descriptive method approach congruent with the V-Model for the software development method approach. And to validate and verify the features of the application ISO9126 was used to evaluate the application based on its functionality, reliability, usability, efficiency and portability which marked a score of 4.54 weighted mean "Very Good".

Keywords: File Management, Sorting, Windows, Files.

Introduction

Technology made great changes [5] as it rapidly became a trend in the society [3]. Along with it, innovation has made possible for the operations in terms of creation, storage, organizing and retrieval of data records [11]. Furthermore, innovation results to larger communications and connections [4] thus, giving birth to at least a million data per minute [16]. As the humanity reaches its information age, every human being digitally processes one-hundred (100) gigabytes of information daily [20]. The produced data doubles in every two years [21]. It is predicted that a single person is estimated to have at least five-thousand two-hundred (5,200) gigabytes of information stored in respective devices and accounts by 2020 [15].

Therefore, the computer market presently offers variety of opportunities of storing huge amount of personal or corporate information in digital form [17]. Storage devices include internal and external hard drives, flash drives, memory cards, and so on [5]. The collection of digital documents, presentations, images, music, videos, databases, and messages are called a digital universe [26] and these are stored in a form which may be space-consuming [23].

Any process that saves space is considered vital in many applications today [22]. Space is an important quantity, which has to be efficiently managed [11]. Wastage of space for unorganized files is always nonnegotiable [9]. To properly store and organize data on media, it uses a method called file system [4]. It is used to control how data is stored and retrieved in the storage medium [2]. The simple principle on where the file is stored is called directories [1]. These directories can contain one or more files or additional directories, branching out into a hierarchal file structure [20]. This organizes the information gathered, produced and stored. With the fast progression of digital data exchange and increased numbers of stored information, developing new approach for efficient sorting, compression and encryption methods are playing a vital role in performance [21]. As more and more information is stored on computers or

communicated via computers, the need to insure that this information is invulnerable to snooping and/or tampering becomes more relevant [25]. Collected, organized and stored data are better if compressed [14]. Compression is a strategy for encoding that permits considerable diminishment in the aggregate number of bits to store or transmit a document [11]. Encryption is the process of scrambling a message so that only the intended recipient can read it as a mean of securing information [12].

A. Project Context

Nowadays, computer technology are mostly focusing on storage space, security and management [23]. Since the beginning of computing, there have always been problems for the largest machines of the day [20]. This situation persists even with powerful CPUs and shared-memory multiprocessors of today [10]. It has become an important issue in modern world as the popularity and infiltration of larger technologies has emerged, making them a prospective medium for disorganization and unsafe keeping [11]. Sorting, Compression and Security is the prerequisite and the most challenging part in any technology [2]. With this, the researchers proposed to develop a desktop application for Windows with an automatic sorting and file organization function that will ease this problem when it comes to rearranging files inside the file system, elimination of file duplication, compressing file to reduce space and password protection.

B. Statement of the Problem

General Problem

The researchers found out that majority of the users are not managing their files properly and the existing functionalities of window file system they are using offers limited functionality which not fully help them in organizing their files.

Specific Problems

1. Majority of the users have difficulty in managing files in their respective personal computer - The researchers conducted one-on-one interview to 20 teaching personnel asking them if they used the existing window file management system to organize their respective files. Based on the results, 3 out of 20 sometimes used, 13 out of 20 rarely used 3 out of 20 every once in a while and 1 out of 20 never used. The result lead the researcher to analyze the situation through verifying the capability of the built in windows file system. The existing system allows to sort files by name, size, item type and date modified in ascending or descending order be done only where the files are located. In addition, the researchers asked if the built-in system greatly help of their file management. According to them, they use it just to easily search a file they are looking for but not really on the aspect of organizing their files as they wanted.

2. The existing built-in windows system allows duplication of files in the file system - Existing windows file system allows the user to have a duplicate copy of the same file with the same name within the same directory but renames it with a “-copy” suffix attached to it. Thus resulted to redundancy of files and consumed storage space

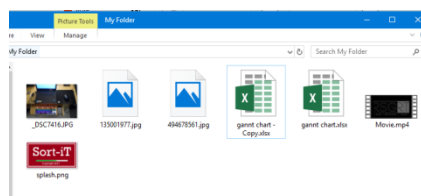


Figure 1: File Duplication

Fig. 1 shows the duplication of file in same directory y “My Folder” has two identical files with the same exact file size listed. This is a common mistake encountered by the user when copying and pasting files from one directory to another.

3. Vulnerability to data theft - The researchers conducted an investigation particularly to teachers and staff that handle confidential files. Existing functionality in the Windows operating system does not allow per-file encryption, and can only achieve it by encrypting the whole disk. The only protection the user has from data theft is their login password that prevents the attacker from logging into the user’s account and take control of all the files inside it. This problem is alarming because according to that failure to do encryption can compromise the security of other information, and the data that is not encrypted with a password is prone to theft because the files are easily accessed even without any confirmation to the owner of the data.

C. Research Objectives

General Objectives

To provide a solution that greatly help the windows users to manage their file through enhancing the functionality of the existing Windows file management system which offers automatic file sorting management by year of creation, file type classification, folder naming category based on its file name content, elimination of file duplication, integration of file compression algorithm and file protection security.

Specific Objectives

1. To design sorting functionality that provides proper file management - The system allows the user to select a specific directory in which he/she will be organizing i.e., "C:/Users/Karl/Desktop/My Folder". By clicking a button in the system, it will automatically start the organization process in which it will create Folders that are based on the files that are found in the chosen directory.
2. To design a checking functionality that checks file duplication during the sorting function - The system checks for duplicate files inside the specified directory during the sorting process. If it found a duplicate file or file with the same name and same type, it will ask the user if he/she wants the file/s to be kept or delete the other one based on the date it was last modified.
3. To design a system that will allow per-file encryption - Specific files chosen from the chosen directory can be encrypted using SHA-256 password encryption that makes it very secure. Files that are encrypted with this method are protected from data theft.

D. Scope and Limitations

Scope

The study focuses on the students of Saint Michael's College of Laguna that use a Windows computer.

1. Graphical User Interface Module - This module is the part that will handle the user inputs in the system and how the user will interact with the system.
2. Sorting Module - This module will handle all the sorting process that the user wishes to do in the chosen directory for the process.
3. Archiving Module - This module will handle the incorporated archiving tool to let the user archive chosen files and folders.
4. Encryption Module - This module will handle the incorporated encryption tool that uses SHA-256 encryption to let the user encrypt previously archived files.

Limitations

The software application will only be dealing with files in a certain directory that the user wants to organize. It will not run on other platforms/operating system such as (OS X, Linux and other Linux Distributions) Information inside files that will be processed will not be altered. The software will only be dealing with a certain directory that the user wants to organize.

E. Significance of the Study

1. Windows users - The study will be useful to the users in managing and organizing files and documents. It will ease the work and lessen the manual process done by the user.
2. Future Researchers - The study will be useful to future researchers for automating the process of sorting the files.

II. Methodology

The researchers used the concept of the V-Model or the Validation Model. The V-Model is where the execution of process happens in a sequential manner in a V-shape. It is an extension of the waterfall model and is based on the association of testing sub-stages for each corresponding development stage.

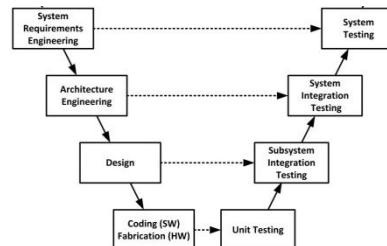


Figure 2: V-Model

2.1 System Requirements Engineering. The researchers conducted observations and small interviews to the needed data and information to deeper understand the requirements and expectation so that the software application meets their specifications. This stage requires extraordinary inclusion from various levels of the association.

2.2 Architecture Engineering. The second developmental stage is to breakdown the requirements into structural format to provide architecture for the needed solution. The system design is prepared. It helps in specifying hardware and system requirements and helps in defining overall system architecture.

2.3 Design. The third development stage is to represent all system processes, inputs, and outputs by developing the graphical user interface. It must have the user-friendliness and convenience for the users.

2.4 Coding. The fourth developmental stage is to process the architecture and design into machine-readable form through programming.

2.5 Testing. The researchers would test of each phase through corresponding testing phases. These are divided in three sub phases: (a) Unit Testing; (b) Integration Testing; and (c) System Testing. In this phase, the developed application will be validated.

This is used to guarantee the overall functionalities of the application. The researchers will use ISO 9126 Software Evaluation to determine the validity of the proposed application.

A. Algorithm

The researchers used two main algorithms to properly execute the main functions for the system, namely: Lossless data Compression, Insertion Sort and SHA-256 bit password Encryption.

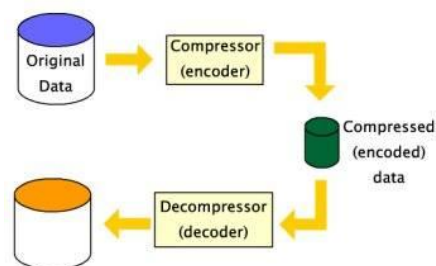


Figure 3: Lossless Data Compression

Lossless data compression is used to archive chosen files in the user-specified directory. This kind of data compression allows the system to shrink down a file into a smaller size without losing quality. Using lossless data compression allows the system to archive and compress files to reduce the space consumed by unused files or old files.

```

i ← 1
while i < length(A)
  j ← i
  while j > 0 and A[j-1] > A[j]
    swap A[j] and A[j-1]
    j ← j - 1
  end while
  i ← i + 1
end while
    
```

Figure 4: Insertion Sort

The researchers used an algorithm that is comparable to insertion sort since it uses an array to store all the information about the files that it will handle during the execution of the sorting process. All the elements of the array will then be rebuilt one-by-one in a sorted in an organized state.

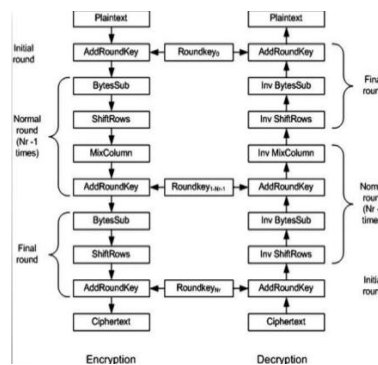


Figure 5: SHA-256 Bit Encryption

AES (acronym of Advanced Encryption Standard) is a symmetric encryption algorithm. AES was designed to be efficient in both hardware and software, and supports a block length of 128 bits and key lengths of 128, 192, and 256 bits. In the designed software the researchers have used the AES-256 encryption. The researchers used AES-256 bit Encryption algorithm because of its efficiency, compared to the 128 bit encryption algorithm the 256 bit algorithm provide more security because of the number of encryption the data will undergo, and when compared to the AES-448 bit algorithm, the 256-bit algorithm has undergone series of attacks and hack testing than the 448 bit algorithm, and the AES-256 bit algorithm has a time proven standard so that is why the researchers used it.

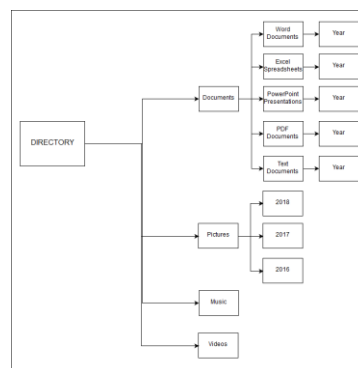


Figure 7: File Hierarchy

Figure 7 shows the developed solution in organizing the user’s files inside the computers. It is solely rely on grouping the sorted files into folders based on their file type attribute i.e., Documents, Pictures, Videos, etc. and then further classifying it to different sub-folders to further distinguish their classification.

III. Results and Discussions

A. Systems Requirements Specifications

The researchers identified the system requirements needed for the furtherance of the development. They are as follows.

Table 1: System Requirements

Windows Operating System Requirements	
Minimum:	Windows 7
Maximum:	Windows 10

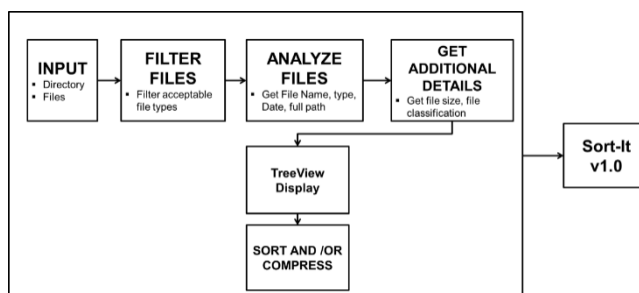


Figure 6: Conceptual Framework

Figure 6 shows the conceptual framework of Sort-It. The process will start by the main input from the user: Directory and Files. The user has to choose and specify the directory that he/she wants to work with. The system will then filter all the acceptable file types that are contained in the chosen directory. The next critical process will be to analyze all accepted files to get the file names, file types, date and its full path. These are the critical information needed to execute proper sorting and organizing of files. Additional details from the files are also obtained. A treeview will then be generated in the system’s UI to display all the files that the system obtained. All the user has to do is to click a button to sort and or compress files.

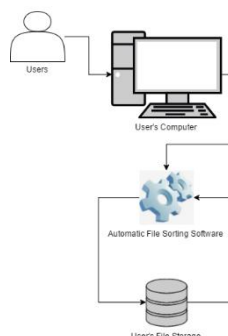


Figure 7: System Architecture

Figure 7 shows the system architecture for the user the can sort the File storage location with the use of Automatic file Sorting software using the user’s own computer.

B. User interface Design

The researchers started the development of the software. The researchers used tools to create the application such as Auto-It V3. The following are the features and user interface design of the application:



Figure 9: Splash Screen

Figure 9 shows the Splash Screen of the user interface of the system when the user opens the Software.

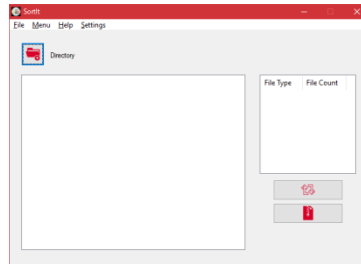


Figure 10: Main Window

Figure 10 shows the main UI for the system. This is where all the information and all actions will be displayed.

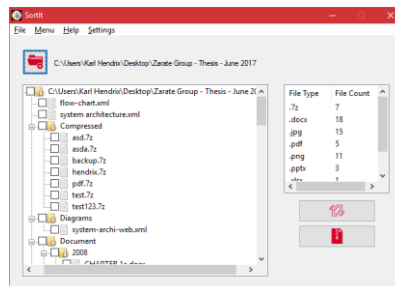


Figure 11: Main Window & its components

Figure 11 shows the display of files contained by the selected folder/directory, it also shows the number of files contained by the selected directory, the folder mark or icon in the TreeView display displays as the folder and the other subfolder contained by the main directory, while the paper icon displays as all the different file types contained by the main folder and sub folders.

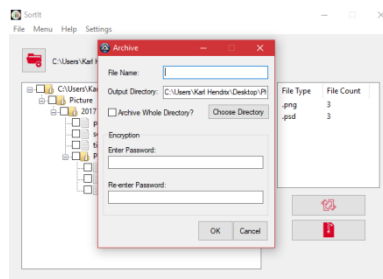


Figure 12: Archive Module & Encryption Module

Figure 12 shows the display after choosing the archive option in the main window, and also the selected folder can be encrypted with a password in the user's own decision.

C. Software Evaluation

Table 2. Software Evaluation

All Characteristics	Mean	Verbal Interpretation
Functionality	4.91	Very Satisfactory
Reliability	4.92	Very Satisfactory
Usability	4.90	Very Satisfactory
Efficiency	4.65	Very Satisfactory
Portability	4.95	Very Satisfactory
Total Weighted Mean	4.87	Very Satisfactory

The researchers submitted the application for software evaluation using the ISO 9126 evaluated by the experts. Based on result the application marked an average weighted mean of 4.87WM, "Very Satisfactory" which means that the majority of the application meet the specified requirements: Functionality marked a weighted mean of 4.91WM, "Very Satisfactory", Reliability marked a weighted

mean of 4.92WM, “Very Satisfactory”, Usability weighted mean of 4.90WM, “Very Satisfactory”, Efficiency weighted mean of 4.65WM, “Very Satisfactory, and Portability with a weighted mean of 4.95WM, “Very Satisfactory”.

IV. Conclusions

The researchers have concluded that the developed system verify enough convenience, sustainability, and ease for the user to successfully manage, sort, or compress files in their computers.

1. Majority of the users have difficulty in managing files in their respective personal computer - The developed software helped ease the difficulty of the users in managing their files in their respective computers.
2. The existing built-in windows system allows duplication of files in the file system - The developed software can detect duplicated files and asks the user to choose what file to delete and what file to keep based on the date the file was last updated.
3. Vulnerability to data theft - The developed software helps the user avoid being the victim of data theft by giving the user option to encrypt files inside their computers.

V. Recommendations

The researchers recommended that the future researchers who want to improve the Automatic File Management Sorting add the following features:

1. Create a version compatible for other platforms such as OS X, Linux and other Linux Distributions.
2. Create a version of system for non-professional users that has a more user-friendly interface.

VI. References

- [1].Apprate, M. (2012). Assisting Manual Sorting of Mail Articles. website: <http://www.google.ch/patents/US6370446>
- [2].Brooek, K.D. (2014). Comparative between Data Compression Algorithm. Adelaide, Australia: University of South Australia.
- [3].Bryan, C. (2014). How does Encryption Work, Is It Really Safe?. Retrieved from Make Use of website: <http://www.makeuseof.com/tag/encryption-care/>
- [4].Echos, T.U. (2013). Compression & Archiving. Retrieved from PeaZip website: http://www.peazip.org/file-compression-performances.html#what_is_data_compression.
- [5].Garg, M. and Sidhu, A.S. (2014). Research Paper on Text Data Compression Algorithm using Hybrid Approach. New Dheli, India: University of New Dheli. Retrieved from: <https://pdfs.semanticscholar.org/94e0/b84eb9b16b1aefc1da377b99c1c425b11466.pdf>
- [6].Gul, M., Amin, N. and Suliman, M. (2015). An Analytical Comparison of Different Sorting Algorithms in Data Structure. Bengaluru, India: Indian Institute of Science.
- [7].Hamel, G. (2013). Advantages and Disadvantages of Traditional File Organization. Retrieved from Chron website: <http://smallbusiness.chron.com/advantages-disadvantages-traditional-file-organization-41400.html>
- [8].Heinz, J. (2014). Reducing Disk Space. Retrieved from Nongnu website: <http://www.nongnu.org/storebackup/en/node21.html>
- [9].Kaur, A. (2013). A Review on Data Compression Techniques. New Dheli, India: Indian Institutes of Technology. Retrieved from: <http://ijcst.com/vol42/2/harpreet.pdf>
- [10]. Keller, N., Shamir, A., et al. (2011). Key Recovery Attacks of Practical Complexity on AES-256 Variants with up to 10 Rounds. Luxembourg: University of Luxembourg. Retrieved from: <https://eprint.iacr.org/2011/374.pdf>
- [11]. Khoo, C., Luyt, B., Eue, C., Yong, S., Osman, J. and Lim, H.H. (2015). Users Organize Electronic Files On Their Workstations In The Office Environment. Singapore: Division of Information Studies, School of Communication & Information, Nanyang Technological University. Retrieved from: <http://www.informationr.net/ir/12-2/paper293.html>
- [12]. Khovratovich, D. and Biryukov, A. (2014). Related-Key Cryptanalysis of the Full AES-192 and AES-256. Luxembourg: University of Luxembourg. Retrieved from: <https://eprint.iacr.org/2014/317.pdf>
- [13]. Laksh, K. & Remzi, P. (2011). An Analysis of Data Corruption in the Storage Stack. Toronto, Canada: University of Toronto. Retrieved from: <http://www.cs.toronto.edu/~bianca/papers/fast08.pdf>

- [14]. Levi, N. (2016). A More Extensible Transparent Encrypted File System Based on Filter Driver. New Jersey, USA: University of New Jersey. Retrieved from: <http://www.jocm.us/uploadfile/2016/0421/20160421031916723.pdf>
- [15]. Ma, S. (2011). Using Hierarchical Folders and Tags for File Management. website: <https://idea.library.drexel.edu/islandora/object/idea:3271>
- [16]. Nardi, B., Anderson, K., and Erickson, T. (2013). Filing and Finding Computer Files. California, USA: Cupertino Publishing. Website: https://www.researchgate.net/publication/228561829_Filing_and_Finding_Computer_Files
- [17]. Porwal, S., Chaudhary, Y., Joshi, J. & Jain, M. (2013). Data Compression Methodologies for Lossless Data and Comparison between Algorithms. Kota, Rajasthan: Rajasthan Technical University. Retrieved from: http://www.ijesit.com/Volume%202/Issue%202/IJESIT201302_23.pdf
- [18]. Schmuck, F. and Haskin, R. (2012). GPFS: A Shared-Disk File System for Large Computing Clusters. Berkley, California, USA: USENIX. Retrieved from: https://www.usenix.org/legacy/events/fast02/full_papers/schmuck/schmuck.pdf
- [19]. Seather, C. and Stoppani, P., Jr. (2014). System of Device Independent File Directories Using a Tag between the Directories and File Descriptors that Migrate with the Files. Montreal, Canada: McGill University. Website: <http://www.freepatentsonline.com/5333315.html>
- [20]. Skibiński, P. and Swacha, J. (2013). Fast and Efficient Log File Compression. Wroclaw, Poland: University of Wroclaw. Retrieved from: <http://ceur-ws.org/Vol-325/paper06.pdf>
- [21]. Soni, M. and Dr. Shukla, N. (2014). Data Compression Techniques in Cloud Computing. Jabalpur, Madhya Pradesh, India: Gyan Ganga College of Technology. Retrieved from: <https://www.ijarce.com/upload/2016/si/SITES-16/IJARCE-SITES%2044.pdf>
- [22]. Sumathi, P. and Karthilkeyan, V.V. (2013). A New Approach for Selection Sorting. Tehran, Iran: Sharif University of Technology. Retrieved from: <http://ijarcet.org/wp-content/uploads/IJARCET-VOL-2-ISSUE-10-2720-2724.pdf>
- [23]. Takahashi, J. and Fukunaga, T. (2012). Differential Fault Analysis on AES with 192 and 256-Bit Keys. Japan: NTT Information Sharing Platform Laboratories Nippon Telegraph and Telephone Corporation. Retrieved from: <https://eprint.iacr.org/2012/023.pdf>
- [24]. Tarundeep, S.S. and Kaur, S.S. (2013). Comparative Analysis Enhancement of Insertion Sort Algorithm. Singapore: IEEE Transactions on Knowledge and Data Engineering. Retrieved from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.278.6654&rep=rep1&type=pdf>
- [25]. Verma, R. and Singh, J. (2015). A Comparative Analysis of Deterministic Sorting Algorithms based on Runtime and Count of Various Operations. Oregon, USA: Western Oregon University. Retrieved from: <http://accentsjournals.org/PaperDirectory/Journal/IJACR/2015/12/6.pdf>
- [26]. Well, D.F. (2014). Encryption. Retrieved from Objs website: <http://www.objs.com/survey/encrypt.html>