Original article
https://doi.org/10.21683/1729-2646-2022-22-1-56-62

Developing a method for recovering data on storage media

Vasily S. Zamolotskikh[1]*, Valentina G. Sidorenko[1]
[1]*Russian University of Transport, Moscow, Russian Federation*
**netaonh11@gmail.com*

Vasily S. Zamolotskikh

Valentina G. Sidorenko

Abstract. Aim. *The development of digital technology brings about the need to digitize data with their subsequent storage on digital media. Regardless of how information is stored, it is of value and its loss may cause harm. There are a number of preventive measures (hardware and logical redundancy of various types) to prevent such loss. Should the preventive measures – due to certain reasons and circumstances – fail to protect the data and access to the latter was lost, it must be recovered in a complete and timely manner. In this context, a need arises for a data recovery algorithm that would take into account the hardware features of today's storage media, their logical structure, as well as the specificity of the stored data.* **Methods.** *There are two approaches to information recovery, i.e., all-purpose and personal. The all-purpose approach involves using a minimal number of programs and tools that work with all items. The personal approach implies a large number of programs and tools that address specific issues associated with the loss of access to information. That enables a faster, higher-quality recovery as compared to the all-purpose approach. Additionally, personal programs are normally cheaper than all-purpose software. All-purpose information recovery tools do not provide quality results when applied to large numbers of failure scenarios. A single utility may not be enough for resolving all issues caused by an incident. A readily available template for obtaining an acceptable result does not exist either. Aside from personal software, there are other alternatives to the all-purpose approach, i.e., the manual data recovery programs and hardware and software systems. In cases of minor logical faults (master boot record corruption) manual data recovery software is used. If a drive is affected by critical hardware issues, hardware and software systems are used.* **Results.** *A method of recovering data on storage media of various types was created. It includes an all-purpose and a personal approaches to information recovery, use of software for manual data recovery, as well as hardware and software systems. The method allows recovering data of popular extensions from common file systems and storage media. Compatibility with RAID arrays of all levels is provided. Programs were selected out of eight sets using the analytic hierarchy process with the priority given to the performance criterion. The method was submitted to a number of tests. Testing involved emulation of incidents associated with the loss of access to data. The cost of eliminating various incidents using the developed methodology is estimated.* **Conclusions.** *Based on the obtained test results, conclusions are set forth regarding the efficiency of the personal approach to information recovery.*

Keywords: *storage medium, data organization, analytic hierarchy process.*

For citation: *Zamolotskikh V.S., Sidorenko V.S. Developing a method for recovering data on storage media. Dependability 2022;1: 56-62. https://doi.org/10.21683/1729-2646-2022-22-1-56-62*

Received on: *21.01.2022* / **Upon revision:** *03.02.2022* / **For printing:** *18.03.2022.*

Introduction

Data is of paramount importance, and in some cases, it is worth more than the media it is stored on. The cost of recovery may also exceed the cost of the storage media. It is therefore very important to properly evaluate the information and determine if it needs to be recovered and whether it is cost-effective to do so. This evaluation, as well as the cause of the loss of access to information, will help choose the method for recovering information on storage media that is described in this paper.

There are several definitions of data recovery. Data recovery is a sequence of actions aimed at retrieving information from a storage medium when such information cannot be read in the conventional way [1].

Data recovery is a sequence of actions, in which damaged or unreadable information is accessed and transferred to another device [2]. The key difference between the two is that the second definition directly refers to the transferring of the recovered data to another (i.e., operable) device.

1. Data organisation on storage media

Developing a data recovery method involves analysing the causes of data loss. Classifying the latter requires taking into consideration the way the data is organised on a medium. Data organisation is examined from two aspects, i.e., hardware-specific and logical.

The hardware-specific data organisation depends on the type of storage medium, its physical component. In hard disk drives (HDDs), it is stored on magnetic plates within the hermetic block [3]. In solid state drives, the data is stored on floating gates that are used in transistors of the same name. Solid state drives include an external interface (in case of peripheral connection to a computer), a controller and memory cells. [1]. In optical storage media, the reading beam passes through the bottom polycarbonate layer, hits the information layer and is reflected. [4, 5].

The hardware-specific data organisation has an effect on the causes of data loss.

1.2. Logical organisation

The logical organisation is common and does not depend on the type of storage medium. It is a hierarchy: disk, partition, directory, file [4]. The first sector of a disk contains the master boot record (MBR) that is required for operating system booting. The partition layout depends on the size of the storage medium. Media larger than 2 TB use global unique identifiers (GUID). The file structure has a main header that stores the file signature. File recovery software uses this record to detect files on a drive.

2. Causes of information loss

The causes of information loss are shown in Fig. 1 [6]. Based on data organisation information, three groups of causes of information loss can be distinguished, i.e., hardware-specific, logical and the human factor (Fig. 2).

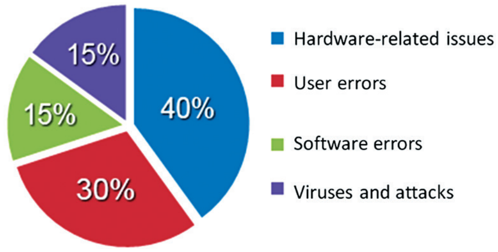

Fig. 1. Distribution of the causes of information loss

3. Algorithm for recovering information on storage media

Based on an analysis of the physical and logical organisation of data on storage media, as well as the causes of data loss, the authors have created an algorithm for data recovery on storage media (Fig. 3). It is based on the use of hardware and logical redundancy in the data organisation, as well as equivalent storage media. Hardware redundancy is implemented by copying service data, backing up and virtualisation, while logical redundancy is implemented by adding a signature to the file header. It is convenient to use operable equivalent storage media for manual data recovery.

The choice of recovery direction depends on the degree of damage to the service information that includes data of the file tables that contain file addresses. If they are severely damaged, the emphasis is placed on the bit sequence analysis, i.e., on signature search. Note that this method is the most popular and is used in most software applications. They, however, also use service information for data recovery. Manual data recovery is done with hex and disk editors. It is convenient to use the disk editor for service data recovery,

Table 1. Baseline data and AHP outputs

Item	SW type	Number of examined SW applications	Winner
1	Storage media testing	5	R.tester
2	Copying media byte-to-byte	5	Active Disk Image
3	Manual data recovery	7	Acronis Disk Director
4	All-purpose	13	Hetman Partition Recovery
5	RAID array recovery	3	RS RAID Retrieve
6	Removed file recovery	4	Active Undelete
7	Corrupted file recovery	12	RS File Repair, Recovery Toolbox
8	Personal	25	Recuva, EasyRecovery

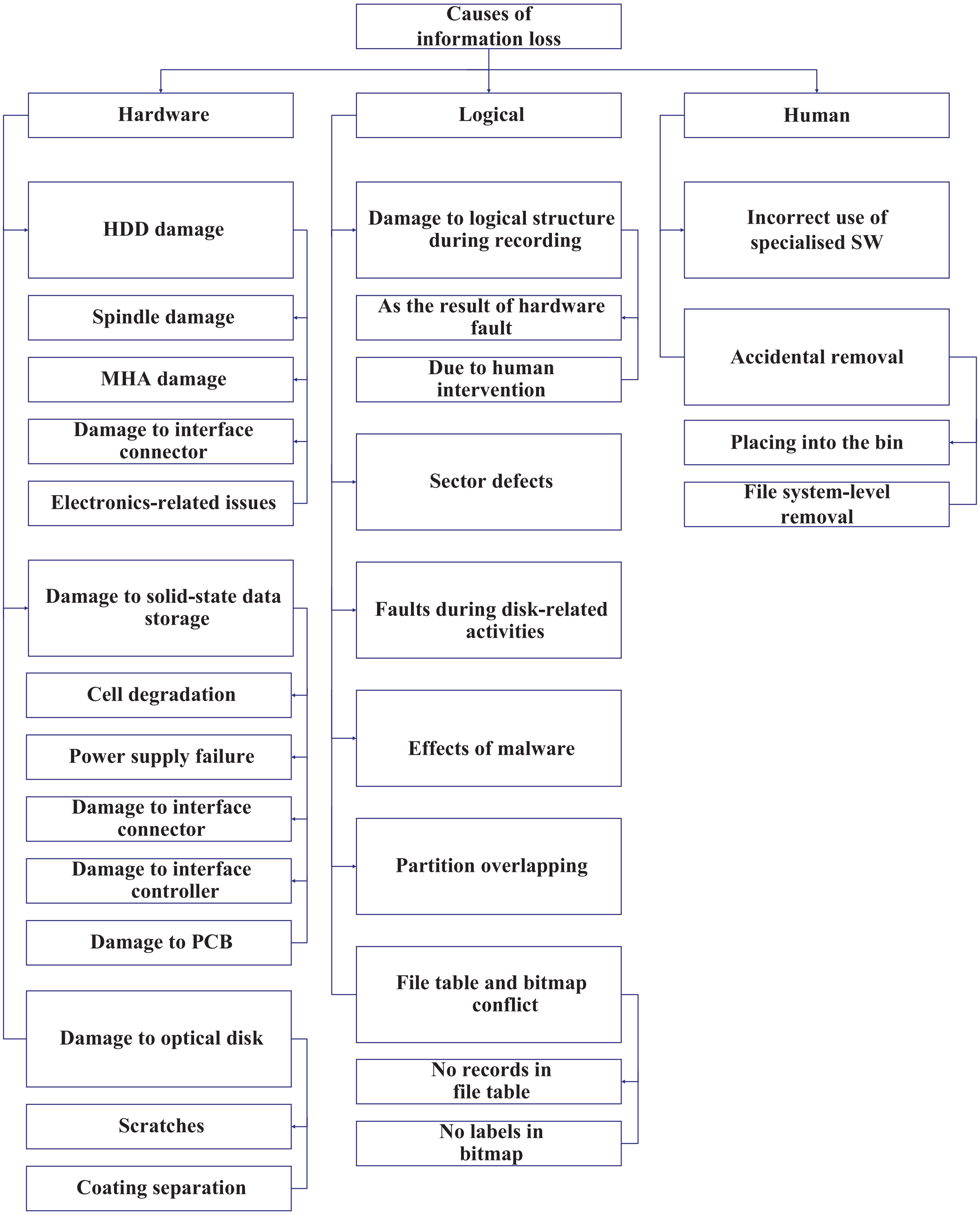

Fig. 2. Causes of information loss

as it represents the storage medium as a partitioned image. Signature search can be done by entering file signatures into the search window [4].

4. Selecting software by means of analytic hierarchy process

For the algorithm to be usable, it must be equipped with software (SW). The SW that is available on the market and can be used for implementing the stages of the developed algorithm was divided into 8 groups. In each group, SW was selected through the analytic hierarchy process (AHP) [7] with priority given to the performance criterion (Table 1).

AHP enables pairwise comparison of items, as well as decision-making based on the item-specific criteria. The choice of this method is due to the following advantages [8]: the possibility of pairwise comparison of criteria, which

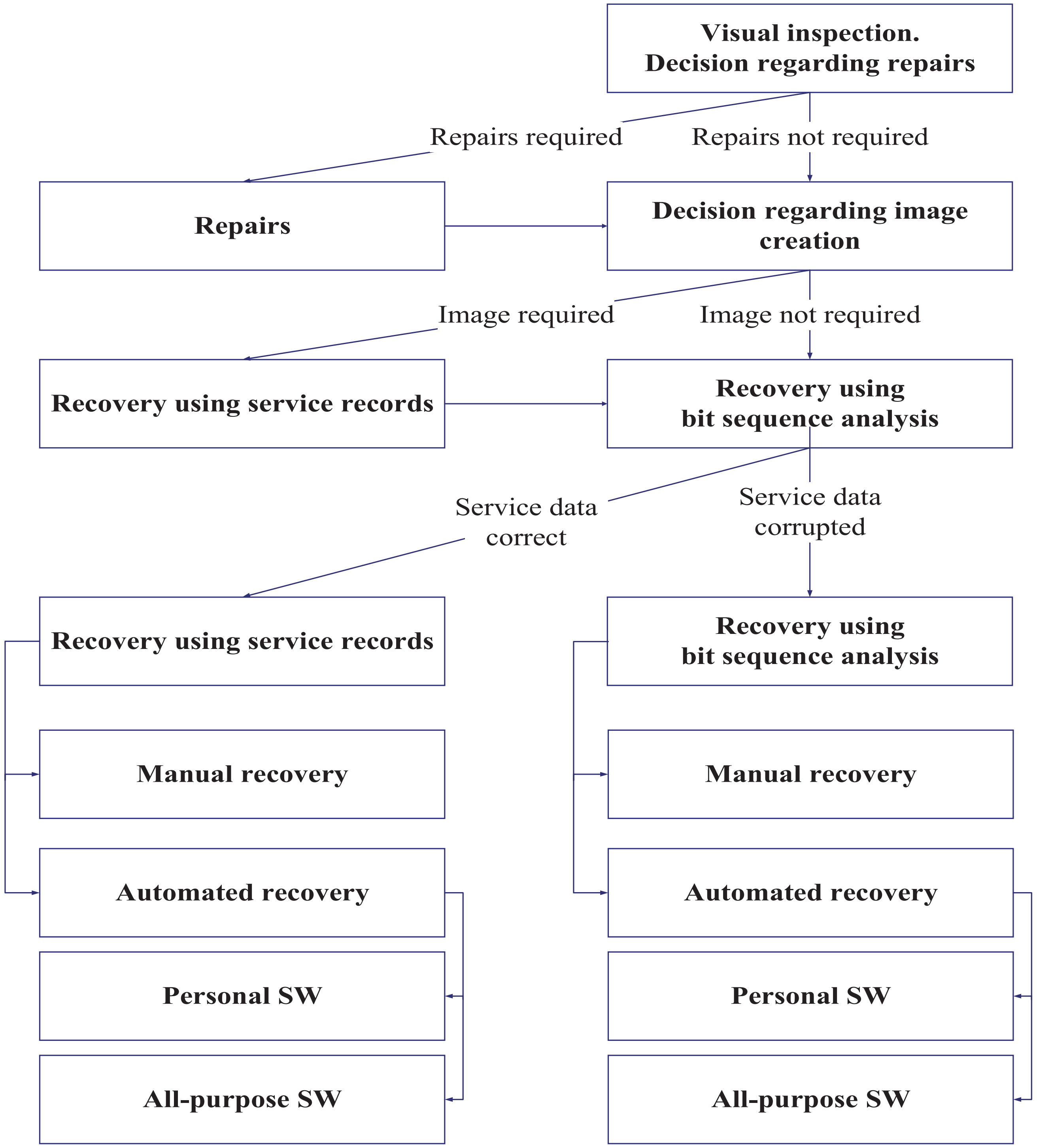

Fig. 3. Information recovery algorithm

facilitates calculations, as there is no need to "keep in sight" all the criteria that can be numerous; the existence of a verbal-numerical scale, which simplifies evaluation, as it compares numbers with easy-to-understand logical comparative constructions; the expert quality assessment (consistency ratio) built into the method.

Formulas (1) to (3) allow estimating items in terms of characteristic criteria.

$$\mathrm{MNGEOM}_x = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n}, \quad (1)$$

where MNGEOM_x is the geometric mean of judgements x (rowwise);

x is a positive number;

n is the number of numbers x.

$$\mathrm{SUM}_a = \sum_{i=1}^{n} a_i, \quad (2)$$

where SUM_a is the result of the sum of the numbers a;

a_i is a number with an ordinal number i;

n is the number of numbers a.

$$\text{VECTOR}_x = \frac{\text{MNGEOM}_x}{\text{SUM}_{\text{MNGEOM}_x}}, \quad (3)$$

where VECTOR_x is the value of the priority vector of criterion/SW x;

MNGEOM_x is the geometric mean of the numbers x;

$\text{SUM}_{\text{MNGEOM}_x}$ is the sum of the geometric mean values of x.

The final estimate will be obtained by using formula 4:

$$\text{GLOBVECTROR}_x = \sum_{i=1}^{n} \text{VECTOR}_{\text{criterion}_i} \times \text{VECTOR}_{\text{SW}_i}, \quad (4)$$

where GLOBVECTROR_x is the value of the global priority vector of SW x;

n is the number of criteria.

$\text{VECTOR}_{\text{criterion}_i}$ is the value of the priority vector of the criterion with serial number i;

$\text{VECTOR}_{\text{SW}_i}$ is the value of the priority vector of SW with serial number i.

The group of formulas from (5) to (7) is required for calculating the conformity relation (CR) that determines the correctness of the judgements. According to the recommendations, it is not supposed to exceed 10%. The recommendation has been fulfilled.

$$\text{NORM}_x = \text{SUM}_x \times \text{VECTOR}_x, \quad (5)$$

where NORM_x is the normalised judgement vector x;

SUM_x is the sum of the judgements values x (columnwise);

where VECTOR_x is the value of the priority vector of criterion/SW x;

$$\text{CI} = \frac{\text{SUM}_{\text{NORM}} - n}{n-1}, \quad (6)$$

where CI is the conformity index of the criterion/SW estimates;

SUM_{NORM} is the sum of the normalised judgement vectors;

n is the number of criteria or SW.

$$\text{CR} = \frac{\text{CI}}{\text{RNCOH}} \times 100\%, \quad (7)$$

where CR is the conformity relation;

CI is the conformity index;

RNCOH is the measure of random coherence.

Table 2. Assessment of AHP criteria

Criterion	Performance	Flexibility in terms of media types	Price	Additional capabilities	Priority vector	Geometric mean	CI	AHP CR, %
Performance	1.0	3.0	4.0	4.0	0.5285	2.6	0.03	3.52
Flexibility in terms of media types	0.3	1.0	2.0	3.0	0.2388	1.2		
Price	0.3	0.5	1.0	2.0	0.1420	0.7		
Additional capabilities	0.3	0.3	0.5	1.0	0.0907	0.5		
Sum	1.8	4.8	7.5	10.0	1	5.0		

Table 3. Evaluation of SW in terms of performance

	Everest Ultimate Edition/AIDA 64	Victoria	MHDD	R.tester	HDDScan	Priority vector	Geometric mean	CI	AHP CR, %
Everest Ultimate Edition/AIDA 64	1.0	2.0	2.0	1.0	2.0	0.2857	1.5	0.00	0.00
Victoria	0.5	1.0	1.0	0.5	1.0	0.1429	0.8		
MHDD	0.5	1.0	1.0	0.5	1.0	0.1429	0.8		
R.tester	1.0	2.0	2.0	1.0	2.0	0.2857	1.5		
HDDScan	0.5	1.0	1.0	0.5	1.0	0.1429	0.8		
Sum	3.5	7.0	7.0	3.5	7.0	1.0000	5.3		

Table 4. Incident models

No. of incident	Storage medium	File system	Cause of incident	Incident manifestation
1	HDD	*NTFS*	HDD impact	MHA damage
2	*SSD*	*NTFS*	Electrical breakdown	Controller damage
3	Flash disc	*NTFS*	*MBR* removal (values replaced with 0)	Flash disc not detected
4	Flash disc	*NTFS*	File deletion at the FS level	Files unavailable
5	Flash disc	*NTFS*	Values of main file headers replaced with 0	Files unreadable
6	HDD	*NTFS*	*MFT* not matching the bitmap (values corresponding to file addresses in *MFT* replaced with 0)	Files unavailable
7	*SSD*	*APFS*	Values replaced with 0: container boot block, disc headers, as well as file removal at the FS level	Drive recognized incorrectly

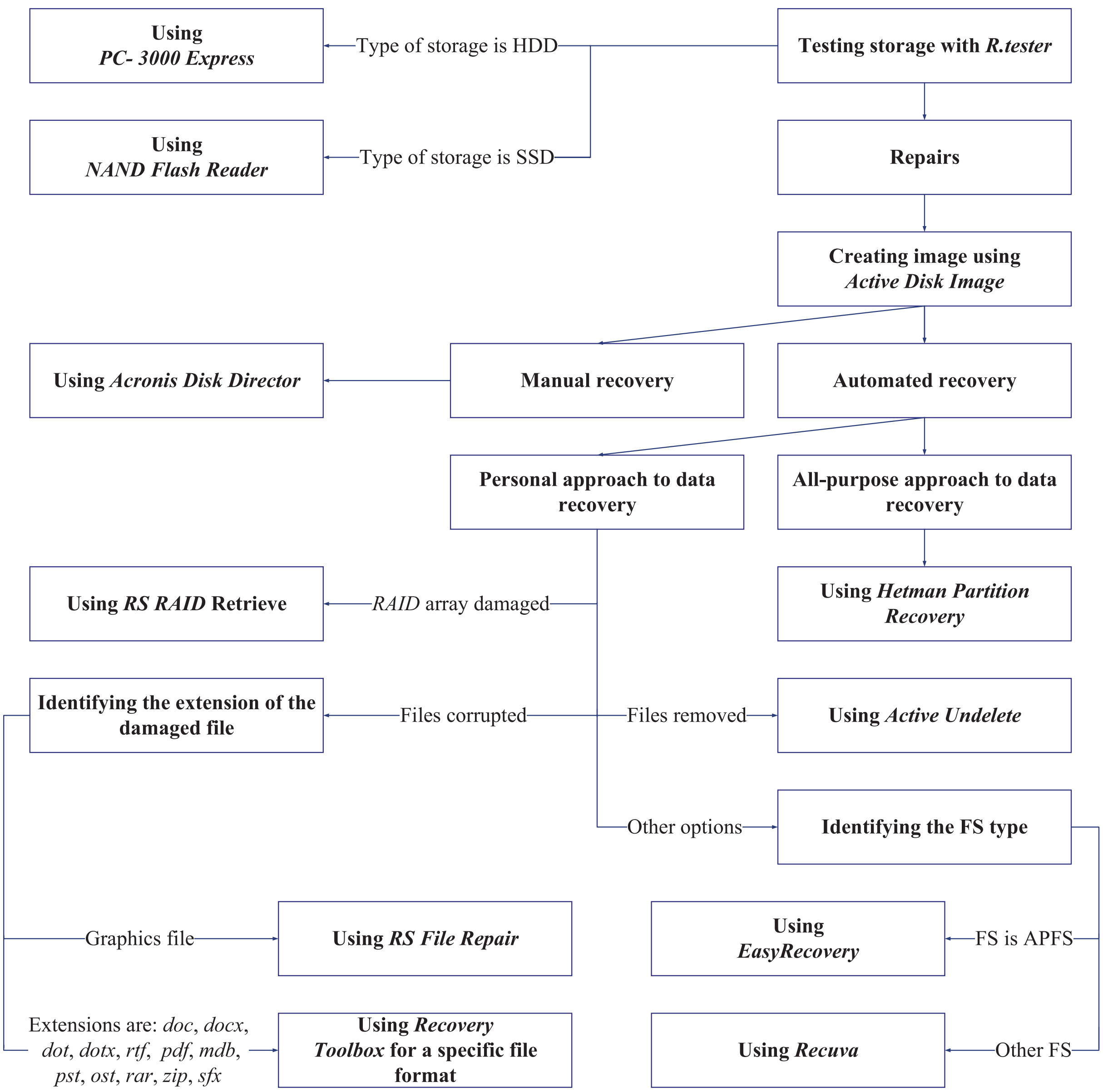

Fig. 4. Method for information recovery

Let us examine the essence of AHP using the case of SW selection for media testing. First, the criteria specific to the test software are evaluated (Table 2). Then, the programs are evaluated in terms of each of the criteria, e.g., the performance (Table 3).

5. Data recovery method

The result is the following method for data recovery on storage media (Fig. 4).

A separate group includes hardware and software systems that should be used if the media are affected by major hardware issues. Manual data recovery software is to be used in case of minor logical faults. An all-purpose approach to data recovery is implemented using *Hetman Partition Recovery*, while a personal approach is implemented using a combination of software at the bottom of Fig. 4.

6. Testing the method

The developed method was tested by solving data access loss incidents (Table 4). The first and second incidents involve the use of hardware and software systems, the third incident involves the use of software for manual information recovery. In the fourth to the sixth incidents, personal information recovery software was used, while the seventh incident involved the use of general-purpose software.

All incidents were successfully resolved. Personal software provided the quickest results, which gives reasons to recommend it for data recovery in most cases.

Table 5 shows the economic models that reflect the options for applying the data recovery method to the incidents, on which the developed method was tested. The cost of technical equipment includes the cost of equipment and software. The cost of operation depends on the consumed time and electrical power. The rapid operation of personal

Table 5. Economic models of the applications of the developed data recovery method

No. of economic model	No. of incident	Used tools	Cost of equipment, RUB	Cost of operation, RUB
1	1	R.tester, PC-3000 Express	191 900	4841.58
2	2	R.tester, NAND Flash Reader	238 243.18	4024.94
3	3	R.tester, Active Disc Image, Acronis Disk Director	54 709.82	8494.19
4	-	R.tester, Active Disc Image, RS RAID Retrieve	59 318.13	3298.1
5	4	R.tester, Active Disc Image, Active Undelete	53 264.11	4068.11
6	5	R.tester, Active Disc Image, RS File Repair	52 938.13	2869.55
7	5	R.tester, Active Disc Image, Recovery Toolbox	53 770.84	2869.55
8	-	R.tester, Active Disc Image, EasyRecovery	57 115.37	2896.34
9	6	R.tester, Active Disc Image, Recuva	52 619.13	3298.1
10	7	R.tester, Active Disc Image, Hetman Partition Recovery	54 818.13	4101.62

software (economic models 4 to 9) ensures monetary savings in operation.

Conclusion

Based on the conducted activities and research, a method for recovering information on storage media was developed (see Fig. 4).

According to test results and economic model calculations, the authors recommend using personal information recovery software.

References

1. [Dictionaries and encyclopaedia on academic.ru. Data recovery]. [accessed: 02.09.2021]. Available at: https://dic.academic.ru/dic.nsf/ruwiki/1369930. (in Russ.)
2. [Storelab. What is data recovery?]. [accessed: 02.09.2021]. Available at: https://storelab-rc.ru/vosstanovlenie-dannih.html. (in Russ.)
3. [How a hard drive works. What is inside an HDD and SSHD]. [accessed: 13.09.2021]. Available at: ttps://www.youtube.com/watch?v=77Ugghti1b0. (in Russ.).
4. Senkevich G.E. [The art of data recovery]. Saint Petersburg: BHV-Petersburg; 2011. (in Russ.)
5. [L-pro. DVD]. [accessed: 07.09.2021]. Available at: http://l-pro.com/uslugi/tirazhirovanie-cd-dvd-diskov/disk-formata-dvd. (in Russ.).
6. [Loss of information leads to loss of business]. [accessed: 26.11.2021]. Available at: https://www.cnews.ru/reviews/itinfrastruktura_predpriyatiya_2013/interviews/aleksandr_frolikov. (in Russ.).
7. Saaty T. Decision making – the Analytic Hierarchy and Network Processes. Moscow: Radio i sviaz; 1989.
8. Korobov V.B., Tutygin A.G. [Advantages and disadvantages of the analytic hierarchy process]. *Izvestia: Herzen University Journal of Humanities & Sciences* 2010;122:108-115. [accessed: 20.01.2022]. Available at: https://cyberleninka.ru/article/n/preimuschestva-i-nedostatki-metoda-analiza-ierarhiy. (in Russ.).

About the authors

Vasily S. Zamolotskikh, Student, Russian University of Transport, Moscow, Russian Federation, e-mail: netaonh11@gmail.com.

Valentina G. Sidorenko, Doctor of Engineering, Professor, Chair Professor, Department of Management and Protection of Information, Russian University of Transport, Moscow, Russian Federation, e-mail: valenfalk@mail.ru.

The authors' contribution

Zamolotskikh V.S. Development of the method for recovering information from storage media, preparation of test models of incidents, testing and analysis of findings, conclusions.

Sidorenko V.G. Analysis of methods and principles of information recovery, overview of the methods for recovering information on storage media.

Conflict of interests

The authors declare the absence of a conflict of interests.