

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.Doi Number

Validation of the authenticity of printed documents (May 2023)

António Ramos¹

¹Universidade de Aveiro, Aveiro, Portugal

Corresponding author: First A. Author (e-mail: ajframos@ua.pt).

ABSTRACT Nowadays, there's a need to protect certain digital documents to avoid information leaks. One of the ways that can be managed is by using a watermark. A watermark inputs information, either in a visible way or not, on a document to guarantee its originality and authenticity. In this dissertation, methods will be approached and analyzed and one developed and implemented for a company.

INDEX TERMS barcodes, digital documents, integrity, printed documents, security, watermark

I. INTRODUCTION

As part of the Master Thesis in Computer Science at the University of Aveiro, it is intended to develop an application module to be integrated with a classified information management solution developed by iCreate Consulting. This solution allows for the printing of classified documents, and mechanisms must be implemented to mark them to ensure their authenticity.

II. DESCRIPTION OF EXISTING SYSTEM

These documents are handled in a physical building with heavy security, where each employee must be reputable. The access to a classified document must be reviewed periodically to ensure the position of the person.

Each employee has access to an assigned job location, and to the documents that are available to the position.

On the other hand, that user only has permission to see documents up to the level associated with, for instance, if they are attributed a certain level of confidentiality, they will only be able to check until that level.

III. GOALS

To secure the authenticity on documents it's two methods are proposed.

The first is creating a barcode that grants validation in an explicit way if a document had origin in the system where the metadata of the document is located. The idea is passing a barcode reader and comparing the information on the barcode if is the same as the document, using database with ids.

The last one is generating a unique watermark, that has the information of text of a document, for instance the words or certain characters.

IV. PROPOSE SOLUTION

The proposed solution is divided into two different actions, process and rectification of document.

In the process option, the goal is to put a watermark in the file that contains information about metadata or text, for that a barcode 128 is used in order to achieve the information about a file that has a id linking to a database.

Rectification is a process that compares the information obtained by the barcode and the document. For instance, if the barcode that decodes the information about a document is assigned by user, and the file being read has another user associated that means if they are attributed a certain level of confidentiality, they will only be able to check until that level.

The proposed algorithm handles 3 types of documents, original, with watermark, and scan file with watermark.

The user is responsible for choosing the action required (process and rectification), accept or reject the document with a watermark that was generated, compare the information in the document with the information in the barcode, and check the document integrity (by comparing characters in the database with the document integrity).

V. Process

Processing a file consists in reading metadata and the content of the file, where metadata represents characteristics of a document like register number, exemplar number, number of copies, classification, state of the document, format, user and date of operation, and content is the text inside of the document.

Based on this information the barcodes 128 and 39 are generated and inserted in footer and header respectively.

When the user approves the new document with watermark, the information is inserted to the database for rectification proposes.

VI. Rectification

To review a document, the document must contain a watermark and these are in the database.

The program is going to try decoding barcode 128 based on footer, and if it succeeds, the program will show the information of the metadata in the database to the user (figure 1).

If the information doesn't appear to be equal, then the user can click to verify the integrity.



Figure 1 - Rectification document

VII. Integrity

Before explaining the integrity of the algorithm there must be an introduction to the concept entropy.

The entropy of document consists in evaluating how good a document is protected. The bigger entropy the bigger the security of a document. To achieve that it's necessary to add the random component in the algorithm.

The algorithm consists in reading the position of barcode 128 in the footer and calculate 9 positions on the documents, where in these 9, 3 are going to be added randomly. In these there are going to be created 3-line segments, equivalent to 81 points ($9 \times (3+3+3)$).

The main objective of using line segments is to calculate the intersection between them, and check if the point is on the letter of the document.

VIII. Example

To test the algorithm, a random document is created with the following text (figure 2). The algorithm processes the document inserting the barcode 39 (header) and barcode 128 (footer), due to the dimensions of the file, it can't show all the document.

A possibility of the integrity resulted by the process is shown in figure 3. The text is in Portuguese but it also works with English, what matters is that the letters are coded in utf-8 and Latin.

For testing the integrity of the document, the word "importante" is replaced, where "n" is associated in database, to "necessário" (figure 4 – black arrow) demonstrating that the algorithm can detect replacing in that zone.

For testing the alteration of the text by removing one paragraph (figure 5), resulting in changing of the points in y axis.

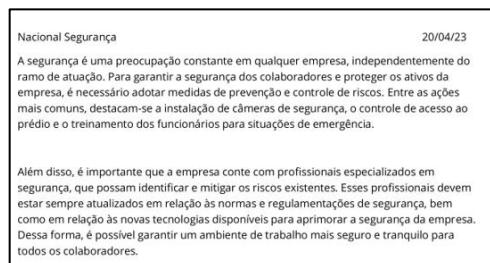


Figure 2 - Original document

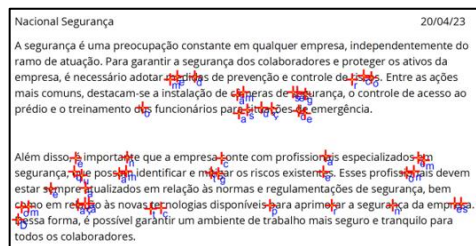


Figure 3 - Integrity

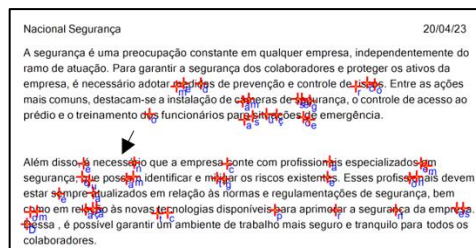


Figure 4 - Replace word

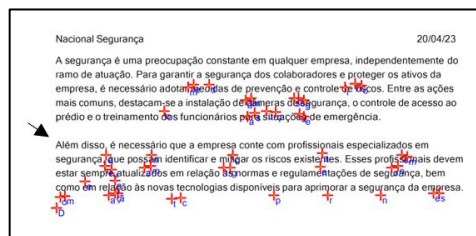


Figure 5 - Remove paragraph

IX. Conclusion

The algorithm developed satisfies the initial objectives of the creation of a barcode to quickly validate a document and create a watermark capable of guaranteeing that document is secure against alteration. The algorithm only works for the first page of large documents, if need it, can be upgraded by inserting barcodes and reading all the file, this will increase the time of processing.



computer science

ANTÓNIO RAMOS was born in Viseu, Portugal. In 2020 graduated in Computer Science at IPV, Viseu, Portugal. From 2020 to 2023 – master's degree on