Safe-PDF-Redaction-Tool

Submitted on: **November 1, 2023**

Lennaert Feijtes
lennaert.feijtes@student.uva.nl
University of Amsterdam
Amsterdam, The Netherlands

Maarten Marx
maartenmarx@uva.nl
University of Amsterdam
Amsterdam, The Netherlands

1 INTRODUCTION

Documents may contain sensitive personal information which can not be shared when, for example, communication between two parties are made publicly available. It is important that this information is left out of the document before being shared. Text is often redacted by masking text with black boxes or by blurring which would ensure confidentiality.. However, this is not always the case [4].

Text redaction tools are essential in various domains. These domains include legal, healthcare, government and corporate sectors where privacy and data protection are of paramount importance. Text redaction tools are used to safeguard sensitive information by hiding or obliterating portions of text, making it unreadable or unintelligible while preserving the structure of the document. No reference to the redacted text should be left in the document after the redaction and no means of identification should be possible by its relation to the non-redacted text i.e. subpixel-sized horizontal shifts that can be recovered from both the redacted and non-redacted characters.

This research is situated in the field of text redaction tools. In this field there is a focus on developing and improving methods that enable the secure and efficient redaction of text in digital documents, in this case PDF. Safe text redaction ensures confidentiality and the safeguarding of sensitive personal information, making this research a significant contribution to the field of text redaction.

Overall, this research deals with the intersection of information security, document processing and algorithmic techniques.

2 RELATED WORK

This research will build upon a great amount of existing literature within the field of computer science, specifically in the domain of text redaction techniques, safety concerns and automation.

Hidden information. A lot of work has previously been done on identifying security concerns related to PDF documents and instances where the security and redaction of Personal Identifiable Information (PII) has been compromised. Hidden information found in the document's version history, track changes, metadata and revision recovery are potential security compromises which can leak confidential information [9] [5] [6].

Mosaicing and blurring. Redaction through mosaicing and blurring, where text is significantly distorted and transformed such that it is unrecognizable for the eye, have been proven to not be viable techniques for text redaction. Due to predictable regularities in text enough information may remain to narrow down the possibilities or even to recover the redacted text [8].

Masking. Redaction by masking text with a so called 'black box' has been proven to also have confidentiality issues related to both the correct implementation as to the remaining information after redaction. Many examples show that the black-box-redaction is prone to mistakes. Often text is only visually hidden or made illegible, but is not actually removed from the original document [1] [3]. Furthermore, recent research has proven that redaction is broken by subpixel-sized horizontal shifts that can be recovered from both the redacted and non-redacted characters which can be used to effectively deredact first and last names [3]. These findings affect also redactions where the text underneath the black box is removed from the document.

3 RESEARCH QUESTION

Research Question. Many security concerns within the PDF text redaction field exist and often confidentiality and privacy are compromised as a result of either human error or information left behind in the document. In my research I want to answer the following: *How can we design and develop a PDF text redaction tool that effectively addresses security concerns, including the prevention of information leakage through subpixel-sized horizontal shifts, while preserving non-redacted text.*

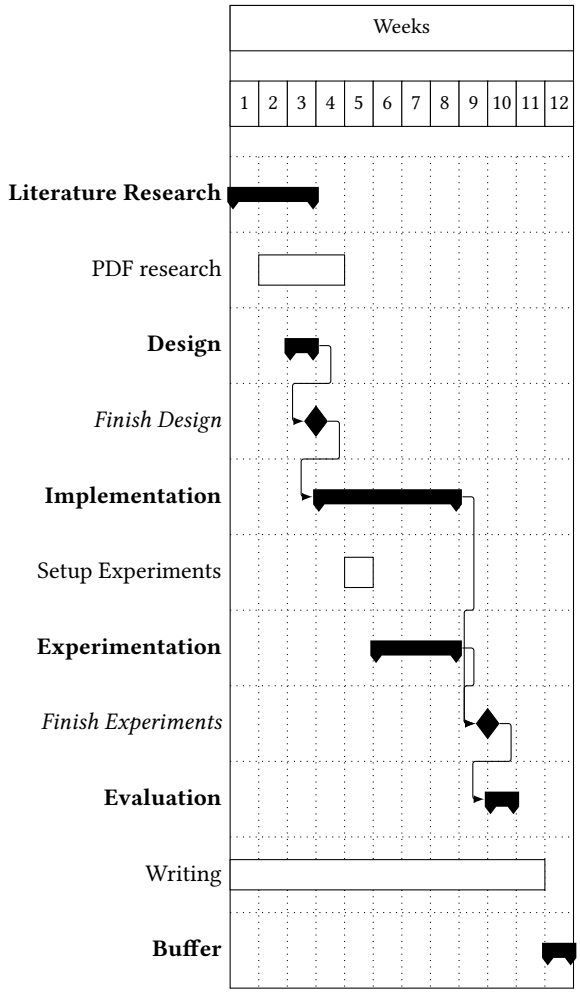
The result. The project will deliver a PDF text redaction method based on research of relevant literature. The redaction method will be tested while designing it using test documents created by myself as well as real-world examples. Relevant tools will be used to assess the safety of the redactions produced by the method in the test cases. *X-Ray Bad Redaction Detector*[10] will be used to check for non-excising redactions where redacted text is retained in the PDF and only visually obscured and the *Edact-Ray Tool Suite*[3] to locate both non-excising and excising redactions. For the latter the code is not publicly available, but for scientific purposes and upon request, the author is willing to share the full code and the dataset [2]. To check metadata, *ExifTool* seems like a promising candidate [7]. Finally libraries such as *pdfid*, *pdf-parser* and *peepdf* seem useful.

4 METHODOLOGY

An extensive review of the literature is needed to gather information about the PDF file extension; manipulation, readability, structure etc. Meanwhile, further research into the security concerns, redaction techniques and legality needs to be done to gain an even better understanding of the problem and the efforts made in the field. What follows is a hands on and practical approach where gained knowledge is put into action and experimentation. First, I have to create the right environment to conduct design, implementing and testing; gather the necessary tools, examples and code. A design of a safe PDF text redaction tool has to be created and implemented. This process must be executed incrementally, with

design and testing proceeding simultaneously to ensure adherence to safety requirements.

5 PROJECT PLAN



REFERENCES

[1] National Security Agency. 2005. Redacting with Confidence: How to Safely Publish Sanitized Reports Converted from Word to PDF. https://sgp.fas.org/othergov/dod/nsa-redact.pdf

[2] Maxwell Blan. 2021. *Story Beyond the Eye: Glyph Positions Break PDF Redaction paper. Github and code.* https://github.com/maxwell-bland/deredaction

[3] Maxwell Bland, Anushya Iyer, and Kirill Levchenko. 2022. Story Beyond the Eye: Glyph Positions Break PDF Text Redaction. *arXiv preprint arXiv:2206.02285* (2022). https://arxiv.org/abs/2206.02285v3

[4] Herbert Dixon. 2019. Embarrassing Redaction Failures. *American Bar* (2019). https://www.americanbar.org/groups/judicial/publications/judges_journal/2019/spring/embarrassing-redaction-failures/

[5] Jock Forrester and Barry Irwin. 2005. An investigation into unintentional information leakage through electronic publication. *Information Security South Africa* (2005). https://www.researchgate.net/publication/229014289_An_Investigation_into_Unintentional_Information_Leakage_through_Electronic_Publication

[6] Australian Government. 2018. An Examination of the Redaction Functionality of Adobe Acrobat Pro DC 201. https://www.cyber.gov.au/sites/default/files/2023-03/PROTECT%20-%20An%20Examination%20of%20the%20Redaction%20Functionality%20of%20Adobe%20Acrobat%20Pro%20DC%202017%20%28October%202021%29.pdf

[7] Phil Harvey. [n. d.]. *ExifTool. Read, Write and Edit Meta Information!* https://exiftool.org/

[8] Steven Hill, Zhimin Zhou, Lawrence K Saul, and Hovav Shacham. 2016. On the (In) effectiveness of Mosaicing and Blurring as Tools for Document Redaction. *Proc. Priv. Enhancing Technol.* 2016, 4 (2016), 403–417. https://cseweb.ucsd.edu/~saul/papers/pets16-redact.pdf

[9] Jens Müller, Dominik Noß, Christian Mainka, Vladislav Mladenov, and Jörg Schwenk. 2021. Processing Dangerous Paths-On Security and Privacy of the Portable Document Format.. In *NDSS*. https://www.ndss-symposium.org/wp-content/uploads/ndss2021_1B-2_23109_paper.pdf

[10] Free Law Project. 2021. *X-Ray Bad Redaction Detector.* https://github.com/freelawproject/x-ray