Redacting Sensitive Information from the Data

A Synopsis submitted
in Partial Fulfillment of the Requirements
for the Degree of
**Bachelor In Technology
In
Computer Science and Technology**

By
**Prachi Rane (46)
Aishwarya Rao (48)
Diksha Verma (60)**

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY
USHA MITTAL INSTITUTE OF TECHNOLOGY
S.N.D.T Women's University, Mumbai 400049
October 2020

Problem Statement

Redaction is the actual process of removing sensitive information from documents and it needs to be completed across many different industries. It typically involves someone manually going through a document word by word looking for sensitive information like name, address, credit card information etc to remove — which of course is a very time consuming and tedious task that is prone to human error. Within the current context of Information Societies, large amounts of information are daily exchanged and/or released or treated as input data to various AI models. The sensitive nature of much of this information causes a serious privacy threat when documents are uncontrollably made available to untrusted third parties. In such cases, appropriate data protection measures should be undertaken by the responsible organization, especially under the umbrella of current legislation on data privacy. To do so, human experts are usually requested to redact or sanitize document contents which are prone to human error as well as tiresome.

Keywords:
Data privacy; Document redaction; Data sanitization; Information theory; Data semantics

Objective and scope of the project

The overall aim of this project is to identify and remove sensitive information within a document, this is done by blacking out or by removing information that is personally identifiable, sensitive, confidential or classified.

Example: To create a legal document or to delete classified information from a document before it is published.

Our focus is to not only remove the sensitive information from the document but to retain the document's context. In order to achieve this, sensitive words will be identified and will be replaced by their label.

For example, if a name is identified it is removed and the word 'PERSON' is put in its place. This will ensure that the content of the document is still intact and the document will no longer contain any sensitive information.

Methodology

The project is divided into 3 main features:

1. **Auto-Redactor** — Aims to take a user's document, automatically identify and replace the sensitive words and present them with a redacted copy of their document

2. **User Selection redaction** — Aims to take the user's document, automatically identify the sensitive words, and send these words back to the user for them to select which ones they want to be removed.

3. **Labeller** — Aims at making the process of retraining the model to recognize more words easier and quicker.
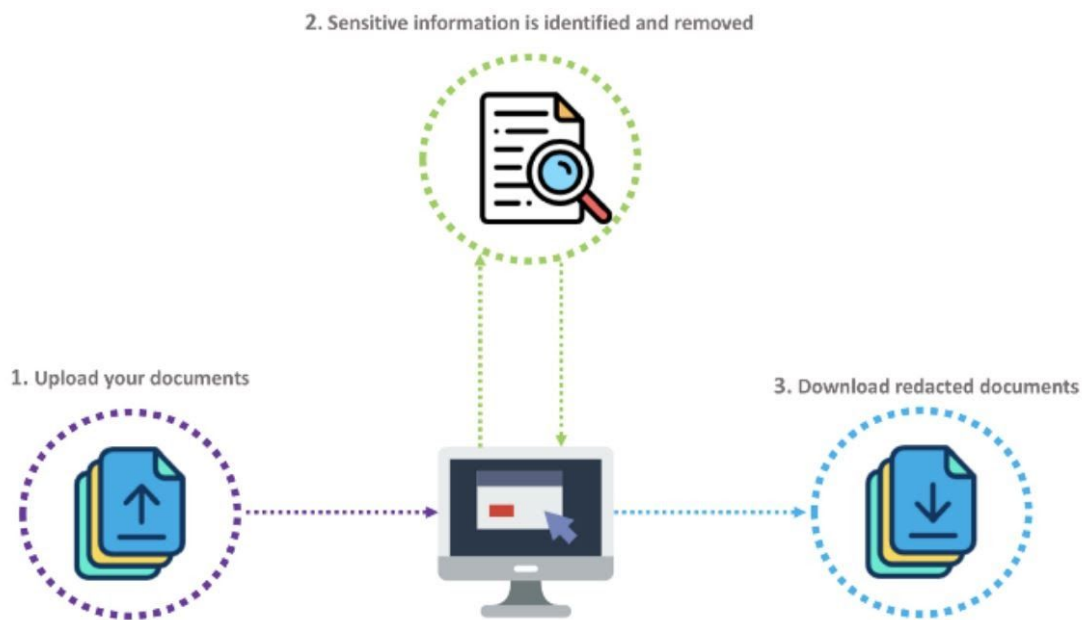
Resources

Computer Languages: Like Python and other open-source libraries of NLP and machine learning.
Software: Anaconda, Spacy, Browser Excel and Jupyter Notebook.

References

[1] Anandan, B., & Clifton, C. (2011). Significance of term relationships on anonymization. Paper presented at the IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, Lyon, France.

[2]Anandan, B., Clifton, C., Jiang, W., Murugesan, M., Pastrana-Camacho, P., & L.Si. (2012). plausibility: Generalizing words to desensitize text. Transactions on Data Privacy, 5,505-534.