Objectives
Accelerate and simplify redaction to protect confidential data in documents and files

Added value for the customer
- Optimization and acceleration of digital redacting processes in documents
- Automated redaction
- Ensuring data protection

Technologies
- Java 1.8
- JBoss Wildfly
- Tesseract OCR (or other)
- NER (Named Entity Recognition)
- Browser: Edge, Chrome, FireFox
- Backend: Operating system neutral

DXC AutoRedact® – Digital automated redaction

Increase the efficiency and effectiveness of redaction in documents and files

Publication Requirements in the EU
Directive 2013/50/EU and amending Directive 2004/109/EC of the European Parliament on the harmonisation of transparency requirements require the redaction of confidential, sensitive and/or personal data.

In addition, the EU Regulation (EU) 2019/1381 on June 20th 2019 on the transparency and sustainability of the EU risk assessment in the food chain enforces the companies that sensitive data must be anonymized, ex.
- Personal Data (Names, Addresses, ..)
- Intellectual Property,
- Patient details must be anonymized as far as possible, e.g., do not mention specific age, ethnicity, or occupation where they are not relevant to the conclusions

Challenges
Due to digital document management and electronic communication, more and more files, documents and data stored electronically have content worthy of protection.

As a result, the efforts to obey legislative requirements are steadily increasing. In the public environment, the redaction of confidential data in documents and files is still done mainly manually. In addition to a high workload of repetitive tasks, the manual process is very fault prone.

Because of the increasing amount of documents and the associated increase in document inspections, this represents a significant cost and time factor to be handled by the authorities in the coming years in addition to the current challenges.

Solution

With the web-based solution DXC AutoRedact®, DXC Technology provides an automated process for redacting confidential data in documents, so that a reconstruction of data is impossible by third parties and thus an optimal support for the data protection is guaranteed.

The goal of DXC Technology has been to create an algorithm and an approach to how typically unstructured information in documents can be automatically detected and classified. However, compared to other solutions, the DXC software offers the added value of automatic pattern recognition, e.g. of phone numbers, names and addresses. This reduces rework done by administrative staff and enables time and cost savings when redacting documents. In addition, a user-friendly interface has been implemented for the most efficient checking of the automated suggestions.

The system is characterized by its self-learning and proactive behavior: Recurring corrections are automatically detected and considered to actively alert the user to problematic areas. For instance, a reference to text passages with poor recognition quality are made if for example, handwritten sections were not captured in full-text recognition.

Functionality

The solution developed by DXC is a browser-based, environment-independent software that offers the following features:

- supports TIFF and PDF documents as input
- Full-text recognition based on semantic analysis and pattern matching (in particular confidential content like names, phone numbers and addresses)
- Interface for efficient review of automated results
- Connection / transfer of documents to a processing (document management) system or E-Akte, or direct application of redaction

The redaction algorithm can also be optimized regarding the environment to increase the hit accuracy:

- Definition of blacklists and whitelists (e.g. do not redact standard addresses of courts)
- For the automated redaction, blacklists can be reused in the same file context.
- Integration and transfer of case data (e.g. plaintiff's name and address, witness name and address, names of judges and prosecutors)

Klaus von Hochmeister
Solution Lead
+49 151 1475 1981
klaus.hochmeister@dxc.com

www.dxc.technology