

# Pseudoanonimizzazione

Roberto Leporini

Università degli Studi di Bergamo

La pseudonimizzazione è una tecnica che consente di oscurare i dati personali in modo tale che le informazioni in un testo non possano essere attribuite a un interessato specifico senza l'utilizzo di informazioni aggiuntive.

Le sentenze fornite dalla Corte d'Appello di Brescia in formato pdf sono state acquisite in formato testo tramite un sistema OCR (scritto in Wolfram notebook) con un'elaborazione batch (in quanto l'ultima pagina di alcune sentenze è un'immagine scansionata).

Tramite espressioni regolari è stato possibile estrarre le informazioni relative ai giudici e alle parti, ma non tutte quelle relative ai testi indicati nella sentenza.

Per questo motivo abbiamo utilizzato delle tecniche di prompt engineering per estrarre direttamente i dati personali dai file pdf.

Usando PDF Replacer abbiamo rimpiazzato i dati con le categorie di appartenenza (e.g., Cognome al posto di Rossi).