

Richtlinie zum „Software-Sprint“

Pseudify 2.0

Schlussbericht

Zuwendungsempfänger:

Ralf Zimmermann

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen **01IS24S24** gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Die Zahl internetbasierter Anwendungen nimmt stetig zu – auch im Zuge fortschreitender Digitalisierung. Viele dieser Anwendungen speichern personenbezogene Daten in großen Datenbanken. Damit Entwickler:innen Fehler analysieren und Funktionen weiterentwickeln können, werden häufig Kopien der Produktionsdatenbanken lokal oder auf Testsystemen bereitgestellt. Dies birgt erhebliche Risiken für die Datensicherheit, da Produktivdaten (z.B. Kundendaten, Adressen, Telefonnummern, Gesundheitsdaten) ungeschützt auf Entwicklungsrechnern liegen können.

Pseudify 2.0 setzt genau hier an: Wie bereits die Vorgängerversion Pseudify 1.0 bietet auch Pseudify 2.0 ein Werkzeug zur pseudonymisierten Bereitstellung von Datenbanken. Allerdings wurde der Funktionsumfang deutlich erweitert, damit auch Anwender:innen ohne tiefgehende technische Kenntnisse davon profitieren können. Kern der Motivation ist es, einer möglichst breiten Nutzerschaft den einfachen Umgang mit pseudonymisierten Daten zu ermöglichen und so das Schutzniveau für personenbezogene Daten zu erhöhen.

Durch eine neu entwickelte, KI-gestützte Analysekomponente erkennt Pseudify 2.0 selbstständig, wo sich personenbezogene Daten in der Datenbank befinden. Außerdem wurde eine benutzerfreundliche grafische Oberfläche (GUI) integriert, die Anwender:innen Schritt für Schritt durch den Pseudonymisierungsprozess führt. Damit wird ein erheblicher Beitrag zur Entlastung von Entwickler:innen und Administrator:innen geleistet, und Pseudify steht gleichzeitig einer erweiterten Zielgruppe offen, die keine oder nur geringe Programmiererfahrung mitbringt.

Beitrag des Projektes zu den Zielen der Förderinitiative „Software-Sprint“

Die Förderinitiative „Software-Sprint“ legt Wert auf gesellschaftlich relevante Lösungen. Pseudify 2.0 leistet hier in mehrfacher Hinsicht einen wichtigen Beitrag:

- **Civic Tech:** Datenschutz ist eine zentrale gesellschaftliche Aufgabe. Indem Pseudify 2.0 die Sicherheit sensibler Daten auch in Entwicklungs- und Testumgebungen gewährleistet, trägt es dazu bei, dass Softwareentwicklung datenschutzkonform und verantwortungsvoll erfolgt.
- **Data Literacy:** Mit der neuen grafischen Oberfläche wird das Verständnis für den Umgang mit personenbezogenen Daten gestärkt. Pseudify 2.0 macht sichtbar, wo personenbezogene Daten liegen und wie diese geschützt werden können. Dies fördert eine bewusste und sachkundige Datenverarbeitung.
- **Gesellschaftlich relevante Lösungen:** Immer mehr Unternehmen und Organisationen müssen den Schutz personenbezogener Daten sicherstellen. Pseudify 2.0 senkt die Hürden für eine verlässliche Umsetzung von Pseudonymisierungsprozessen und erzielt damit eine nachhaltige Wirkung in vielen Branchen.

Ausführliche Darstellung der Ergebnisse

Im Rahmen des Projekts Pseudify 2.0 konnten insbesondere zwei große Ziele umgesetzt werden:

1. Performanceoptimierung

- Die Analyse personenbezogener Daten in großen Datenbanken konnte massiv beschleunigt werden. Eine 5 GB große Testdatenbank, für deren vollständige Prüfung zuvor rund vier Stunden erforderlich waren, wird nun in etwa zwei Minuten analysiert.
- Auch der Pseudonymisierungsprozess wurde stark optimiert. Die Verarbeitung der Testdatenbank (5 GB) konnte von ungefähr 25 Minuten auf nur noch 2 Minuten reduziert werden. Dabei wird unter anderem eine Parallelisierung genutzt, sodass die Ausführungszeiten je nach verfügbarer Hardware weiter skalieren können.

2. Erweiterte Zugänglichkeit

- **Grafische Benutzeroberfläche (GUI):** Wo zuvor Konfigurationsdateien und Kommandozeilen-Interaktionen nötig waren, führt jetzt ein benutzerfreundliches Webinterface durch den gesamten Pseudonymisierungsprozess. Dies erleichtert die Bedienung erheblich – auch für Personen ohne tiefgehende Programmierkenntnisse.
- **KI-gestützte Assistenten:** In Pseudify 2.0 analysiert ein integrierter „Wizard“ die Datenbank automatisch, erkennt potenzielle personenbezogene Daten und schlägt passende

Pseudonymisierungsstrategien vor. Auf Wunsch kann Pseudify daraus vollautomatisch ein passendes Pseudonymisierungsprofil erstellen.

- **Lokale Datensicherheit:** Die KI-gestützten Analysekomponenten arbeiten vollständig lokal, sodass die Daten die eigene Umgebung niemals verlassen. Dies ermöglicht eine rechtskonforme und datenschutzfreundliche Nutzung in sensiblen Umfeldern.

Alle geplanten Meilensteine im Rahmen des geförderten Projekts wurden erfolgreich umgesetzt. Pseudify 2.0 ist weiterhin Open Source und kann von der Community für unterschiedlichste Einsatzbereiche erweitert und angepasst werden.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Zielgruppen von Pseudify 2.0 sind neben Entwickler:innen und Administrator:innen von CI-Systemen nun explizit auch Nutzer:innen, die keine oder nur geringe Programmierkenntnisse besitzen – beispielsweise aus dem Bereich Datenschutz, Qualitätsmanagement, Datenanalyse oder Projektleitung.

- **Nutzen:**
 - **Zeitersparnis:** Durch den vollautomatischen Erkennungsprozess sowie die weitgehend automatisierte Konfiguration sparen selbst erfahrene Entwickler:innen erheblich Zeit.
 - **Rechtssicherheit:** Pseudify 2.0 erleichtert es Organisationen, gesetzliche Anforderungen an den Datenschutz (z.B. DSGVO) zu erfüllen und reduziert das Risiko von Datenschutzverletzungen.
 - **Breite Einsetzbarkeit:** Dank seiner Datenbank- und Applikationsagnostik kann Pseudify 2.0 im Kontext unterschiedlichster Webanwendungen und Branchen eingesetzt werden.
- **Mögliche Weiterentwicklungen:**
 - **Verfeinerung der KI:** Eine stärkere Nutzung von Machine Learning-Methoden könnte die Erkennungsquoten für personenbezogene Daten in komplexen Datenstrukturen noch weiter erhöhen.
 - **Weitere vordefinierte Profile:** Um Einsteiger:innen den Start zu erleichtern, könnten zusätzliche Standardprofile für bekannte Content-Management- und Shopsysteme erstellt und mitgeliefert werden.
 - **Integration in CI-Systeme:** Eine enge Verzahnung mit gängigen CI/CD-Pipelines ist geplant, um Datenbank-Pseudonymisierungen direkt in Automatisierungsprozesse zu integrieren.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Für die Umsetzung von Pseudify 2.0 gab es keine Arbeiten oder Lösungsansätze, die aufgegeben werden mussten. Alle ursprünglich definierten Ziele konnten innerhalb des Förderzeitraums erfolgreich realisiert werden.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Interessierte finden alle Informationen rund um Pseudify 2.0 auf der Projektwebseite <https://www.pseudify.me>. Dort steht eine umfassende Dokumentation zur Verfügung. Der Quellcode ist auf GitHub unter <https://github.com/waldhacker/pseudify-ai> abrufbar. Darüber hinaus sorgen Vorträge auf Entwicklerkonferenzen für Sichtbarkeit.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Die geplante Arbeits- und Kostenstruktur konnte vollständig eingehalten werden. Insbesondere dank umfassender Vorplanung und einer Fokussierung auf die beiden Hauptziele (Performanceoptimierung und GUI-Entwicklung) traten während der Realisierung keine wesentlichen Planänderungen auf. Die Fördermittel wurden zweckgebunden eingesetzt, um Pseudify 2.0 im vorgesehenen Zeit- und Budgetrahmen umzusetzen.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Während des Projektzeitraums traten keine Entwicklungen anderer Projekte oder Institutionen auf, die den Fortschritt oder die Zielsetzung von Pseudify 2.0 wesentlich beeinflusst hätten. Das Projekt hat sich vielmehr als eigenständige Lösung im Bereich der Datenbank-Pseudonymisierung etabliert und konnte unabhängig weiterentwickelt werden.