Richtlinie zum „Software-Sprint“

ACCID – Automated Clustering of Conflict Incident Data

Schlussbericht

Zuwendungsempfänger: Hadi Habal Para Deutch GbR
Abdul Hadi Habal, Jeffrey Deutch, Nicholas Para

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS18S17 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Was war Deine Motivation? Welches Problem wolltest Du mit Deinem Projekt lösen? Wie war die geplante Vorgehensweise zur Problemlösung (auch Angabe der wichtigsten Meilensteine)?

Syrian Archive hat seit 2014 visuelle Dokumentationen über Verstöße gegen die Menschenrechte gesammelt. Die Databenbank dieser visuellen Dokumentationen enthält mehr als 2.000.000 digitale Einheiten, darin inbegriffen sind Videos und Fotos über den Konflikt in Syrien.

Die Datenbank der Syrian Archive bisherr bestehte aus individuellen Einheiten von digitalen Medieninhalten (z.B. Videos und Fotos) anstatt aus kontextualisierten Sammlungen (z.B. Angriffe mit chemischen Waffen) und

machte es daher schwierig für Journalisten und Menschenrechts Forscher zu verstehen, welche digitale Einheit zu welchem Vorfall oder welcher Sammlung gehörte, und ergibt daher zusätzliche Schwierigkeiten beim Verstehen der Größenordnung der Menschenrechtsverletzungen und der Vorkommnisse in Syrien.

Syrian Archive wollte die Finanzierung der Runde 3 dazu benutzen, eine technische und methodische Infrastruktur zu entwickeln, um die digitalen Einheiten in Vorkommnisse zu bündeln - ACCID.

Um ACCID zu entwickeln, wollten wir spezielle tech Lösungen gleichzeitig mit dem Untersuchungen entwickeln. Das bedeutet, dass es keinen genauen technischen Plan gab, sondern dass die Technologie abhängig von den Methodologien und Investigationen gebaut sein sollte.

Durch die Entwicklung von diesen Werkzeugen und Methodologie wollten wir mit den Überprüfungs Team zusammenarbeiten um zu entscheiden, was sie für ihre Arbeit benötigen. Damit wird eine Befehlszeilenschnittstelle gestartet, um Data in suchbare Formate zu leiten und um bündelbare Einheiten für den Benutzer darzustellen.

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

Wer ist die Zielgruppe für Deine Lösung? Wie profitiert sie von den Ergebnissen Deines Projekts? Welche Bezüge gibt es zu den Themenfeldern und Zielen des Software Sprints?

Interessiert an unserer vorgeschlagenen Methodologie und der technischen Infrastruktur sind Journalisten, Menschenrechtsgruppen, Forscher und Akademiker aus Großbritannien, Griechenland, den

Vereinigten Staaten und Deutschland, die zwar an verschiedenen menschenrechtlichen Untersuchungen arbeiten, aber denen ein vollständiger Weg fehlt, um die schon aufbewahrten, kontextualisierten Inhalte zu benutzen. Dazu gehören hauptsächlich: Berkeley Human Rights Center, ein in den US basiertes Forschungsinstitut, wo Studenten in technischer Implementierung, Überprüfung und Methodologien von Bündelungen trainiert werden, Arab Reporters for Investigative Journalism (ARIJ), die daran interessiert sind, eine Archive zu erstellen, bei denen sie die Open Source methoden und die Software benutzen, die von den Syrian Archives in Zusammenarbeit mit anderen Gruppen entwickelt wurden.

Ausführliche Darstellung der Ergebnisse

Welche konkreten Ergebnisse hast Du erzielt? Konnten alle Meilensteine erreicht werden? Welche zusätzlichen Erkenntnisse hast Du aus der Projektarbeit gewonnen, auch im Hinblick auf die Begleitung durch die Open Knowledge Foundation?

Am ende des Projekts ist es einem Prüfer oder einem Ermittler möglich, mit ACCID die Command Line zu benutzen, um die Database zu durchsuchen und sinnvolle Ergebnisse aus dem großen Dataset der gesammelten Einheiten zu finden. Es existiert jetzt eine Methodologie und techtools das Einheiten jetzt auch gebündelt in eine Datenbank gespeichert werden können. Veränderungen in diese daten werden auch gespeichert. Die Methodologie existiert als ein Open Source Aufsatz, und die tech tools als eine sammlung von javascript libraries.

Außerdem ist für das team durch ACCID es jetzt möglich die gebündelte Ereignisse auf Zeitpunkt und Standort zu exportieren, und dies auf eine digitale Landkarte zeigen. Die begleitung der Open Knowledge Foundation hat uns ermöglicht wiederverwendbare technologien zu entdecken, die die entwicklung von ACCID erleichtert hat.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Welcher Nutzen ergibt sich für die Zielgruppe aus den Ergebnissen Deines Projekts? Welche weiter-gehenden Effekte ergeben sich aus der Open-Source-Stellung der Ergebnisse? Gibt es Ideen für die Weiterentwicklung Deiner Lösung und Pläne zu deren Umsetzung?
Hat die Arbeit in dem Projekt Dich in Deiner persönlichen, fachlichen Weiterentwicklung unterstützt?

Durch Datenverdichtung, Konservierung, Katalogisierung und Sicherstellung digitaler Dokumentationen von Menschenrechtsverletzungen in Syrien werden vom Syrian Archive die Syrische Zivilgesellschaft, Syrische Journalisten, MRV (z.B. Human Rights Watch), Organisationen und Rechtsanwälte, die von der Basis berichten, UN- Agenturen, die Menschenrechtsverletzungen dokumentieren (z.B. OHCHR), Forscher, Ermittler und Geschichtswissenschaftler, die Konflikte untersuchen, und Familien von Opfern direkt profitieren.

Die Open Source tools und Methodologie ACCID unterstützen nicht nur Syrian Archive, sondern auch die vorgenannten gruppen und deren zusammenarbeit.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Gab es Arbeiten bzw. Lösungsansätze, die nicht weiter verfolgt wurden? Was waren die Hinter-gründe, und wie bist Du alternativ vorgegangen?

ACCID wurde zusammen mit der arbeit von Prüfer und Ermittlerteams entwickelt. Das bedeutet dass die lösungen direkt fuer die bestehende probleme gebaut wurden.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Wo können sich Interessenten detailliert über Deine Projektergebnisse informieren (z.B. Webseite, GHitHub, Veröffentlichungen)?

Ein exemplar von eine gebuendelte datenbank:
https://syrianarchive.org/en/collections/chemical-weapons/database

Die tech tools um das zu ermöglichen:
https://github.com/syrianarchive/accid

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Gab es im Projektverlauf Ereignisse, die eine Anpassung der Planung erforderlich machten – z.B. Mehr- oder Minderaufwand bei der Bearbeitung von Teilaufgaben?

Nein.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Gab es Entwicklungen anderer Personen oder Institutionen, die Einfluss auf Deine Arbeiten und die Zielsetzung hatten? Wenn ja, worin bestand dieser und wie bist Du damit umgegangen?

Wir haben mit viele vorgenannte gruppen gearbeitet die ähnliche arbeit machen. Wir haben womöglich mit diese gruppen koordiniert und einander unterstützt.