Richtlinie zum „Software-Sprint“

BIGBANG – Ein Softwaretool zur Erforschung soziotechnischer Praxis in der Standardsetzung des Internets

Schlussbericht

Zuwendungsempfänger:

Christoph Becker

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS21S43 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)?

“Infrastruktur setzt die unsichtbaren Regeln, die die Räume unseres Alltags bestimmen”, schreibt die Sozialgeografin Keller Easterling, “Veränderungen in der sich globalisierenden Welt werden nicht in der Sprache des Rechts und der Diplomatie, sondern vielmehr in der Sprache der Infrastruktur geschrieben”. Unsere Motivation für die Entwicklung von BigBang ist, die Machtdynamiken zu analysieren, die Informationsinfrastrukturen erzeugen und damit das Rückgrat der Informationsgesellschaften bilden.

Diskussionen über Normen des Internets beschränken sich allzu oft auf das, was sichtbar ist, z. B. Suchmaschinen und soziale Netzwerke. Die dahinter liegende Infrastruktur, die alles Sichtbare ermöglicht und formt, wird häufig als neutral oder als unveränderlich angesehen. Es ist jedoch auch hier wichtig, die Beteiligten mit ihren Hintergründen kritisch zu untersuchen.

BigBang ermöglicht dies durch die Analyse der Koproduktions- und Wettbewerbsprozessen, Zugehörigkeiten und sozialen Netzwerken, deren Zentralität und Einfluss, Konfliktentstehung und -lösung die sich in öffentlichen Mailinglisten wichtiger Normgremien und Organisationen zur Standardisierung widerspiegeln.

Um diese Problematik anzugehen, ermöglicht BigBang den Zugriff auf und das Herunterladen von Mailinglisten der verantwortlichen Normgremien und Organisationen zur Standardisierung der digitalen Infrastrukturen (z.B. IETF, W3C, RIPE, ICANN, 3GPP, und IEEE).

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?

Zielgruppen kommen vor allem aus der Zivilgesellschaft (z.B. ARTICLE19 und dem "Center for Democracy and Technology"), so wie Forschende der Medienwissenschaft, Computersoziologie, Anthropologie, Politikwissenschaft, des Rechts, und der Wirtschaft. Des weiteren haben Normgremien, so wie die Leitung der “Internet Engineering Task Force” (IETF), Interesse gezeigt um BigBang zu verwenden.

Diese unterschiedlichen Zielgruppen profitieren auf unterschiedliche Weise von unserer Arbeit. Sie lassen sich mit den folgenden Punkten zusammenfassen. Erstens, vereinfache ich den Zugriff und Analyse der öffentlichen Mailinglisten von Normgremien für Menschen die selber nicht die nötigen Programmiersprachen kennen. Zweitens, unser gewähltes Datenformat der Mailinglisten verringert den Datenverkehr im Internet, da beliebige Mailinglisten mit einem klick von Interessenten von uns zu erhalten sind. Drittens, ermöglichen unsere bereitgestellten Funktionen eine Einsicht in das Kommunikationsnetzwerk in den Mailinglisten, die für Forscher qualitative Studien quantitativ Untermauern kann.

BigBang bezieht sich somit vor allem auf die Thematik der Transparenz und Anwendung, da das Sichtbar machen von Dominanz in Normgremien (z.B. von Huawei, Ericsson, oder Samsung) das Ziel ist.

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?

Während der Förderphase habe ich das Herunterladen der Mailinglisten mit den folgenden die Funktionen zum erweitert. Erstens, Mailinglisten können nun automatisch vervollständigt werden durch das gezielte Herunterladen fehlender Emails. Zweitens, können nun auch Dokumente die im Anhang von Emails vorhanden sind Heruntergeladen werden. Drittens, können Mailinglisten gezielter Heruntergeladen werden durch das beschränken auf bestimmte Zeitabschnitte oder Themen.

Des weiteren habe ich die Analyse der Mailinglisten mit den folgenden Funktionen erweitert. Erstens, Mailinglisten können nun schnell und effizient nach beliebigen Informationen die in der Kopfzeilen Metadaten (so wie Sender, Empfänger, Zeitstempel, Betreffzeile) vorhanden sind gefiltert werden. Zweitens, lassen sich gewichtete Graphen der Kommunikationsnetwerke erstellen und nach verschiedensten Zentralitäten untersuchen. Drittens, können nun Ergebnisse (so wie Zentralität, Anzahl an gesendeten oder empfangen Nachrichten, etc.) in beliebige Perioden aggregiert werden. Viertens, habe ich es nun auch ermöglicht Zusammenhänge und Unterschiede des Austauschs zwischen verschieden Marktsektoren und Ländern zu untersuchen.

Darüber hinaus, habe ich BigBang mit den folgenden Fortschritten verbessert. Erstens, eine erneuerte und verbesserte Dokumentation. Zweitens, das erstellen eines Servers auf dem alle Datensätze sicher und einfach zugreifbar gespeichert sind. Drittens, haben wir den Prozess in Gang gesetzt um zusammen mit Juristen von Covington & Burling LLP zu klären wie ich unsere Dareinsetzte rechtlich richtig teilen können.

Es wurden zahlreiche Erkenntnisse durch die Arbeit gewonnen die in Workshops und Konferenzen vorgestellt wurden und werden.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Welcher Nutzen ergibt sich für die Zielgruppe aus den Ergebnissen Deines Projekts? Welche weitergehenden Effekte ergeben sich aus der Open-Source-Stellung der Ergebnisse? Gibt es Ideen für die Weiterentwicklung Deiner Lösung und Pläne zu deren Umsetzung?

Die Zielgruppe und wie sie unsere Arbeit nutzen können wurde unter dem Abschnitt 'Beitrag des Projektes zu den Zielen der Förderinitative „Software-Sprint"' geschildert.

Die weitergehenden Effekte der Open-Source-Stellung unserer Ergebnisse sind vielschichtig. Ersten ermöglicht es Transparenz unserer Methodik die vor allem für Wissenschaftliche Arbeiten wichtig ist. Somit können andere die Veröffentlichten Resultate immer reproduzieren und kontrollieren. Zweitens, kann jeder sich an uns wenden mit detaillierten Vorschlägen wie man das Projekt verbessern und erweitern kann. Drittens, kann jeder BigBang für eigene Zwecke und Fragestellungen verwenden, was insbesondere Aktivisten in der Zivilgesellschaft, Non-Profit-Organisationen, und Forschungsprojekt unterstützt.

BigBang soll in mehreren Richtungen erweitert werden. Ersten, wird ein Fachkurs an der Universität von Amsterdam entwickelt welches BigBang mit einbeziehen wird. Dafür werde ich eine Online-Umgebung erstellen in der man BigBang problemlos vom Webbrowser laufen lassen kann. Um mehrere Mailinglisten zugleich analysieren zu können, möchte ich die parallele Verarbeitung ermöglichen die alle Rechenkapazitäten eines Computers ausschöpfen können. Des weiteren arbeiten ich an einer Veröffentlichung für das 'Journal of Open Source Software' Journal.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

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

Ein Ziel war es um die Mailinglisten in einem Wissensnetzwerk (im Engl. 'Knowledge Graph') darstellen zu können um. Jedoch musste ich feststellen das die Art in der Information in Mailinglisten niedergeschrieben ist, zu vielseitig und dadurch unstrukturiert ist, um jede Email nach Thema und Wissensinhalt akkurat zu Klassifizieren. Dadurch musste ich dieses Vorhaben erst einmal einstellen.

Ich wollte auch alle heruntergeladen Mailinglisten einfach zugreizbar machen und haben erst Git LFS probiert. Jedoch wurde uns schnell klar das unsere Dateien zu gross sind um sie Kostenfrei mithilfe von Git zu speichern. Ein Ausweg habe ich gefunden in dem ich auf einem eigenen Rechner Platz geschaffen haben, auf dem ich und andere mithilfe von SSH zugreifen können. Ich arbeiten an einem Verfahren um es zu ermöglichen noch weiteren die Daten zur verfügung zu stellen.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

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

Interessierte könne sich anhand von GitHub (https://github.com/datactive/bigbang) , ReadTheDocs (https://bigbang-py.readthedocs.io/en/latest/), und einer eigenen Webseite (http://datactive.github.io/bigbang/) über das Projekt informieren. Des weiteren gibt es mehrere Veröffentlichungen die den Nutzen von BigBang motivieren und darstellen, unter anderem:

- ten Oever, N., Milan, S., & Beraldo, D. (2020). Studying Discourse in Internet Governance through Mailing-list Analysis. In D. L. Cogburn, L. DeNardis, N. S. Levinson, & F. Musiani (Eds.), Research Methods in Internet Governance. Cambridge, MA: MIT Press.
- Doty, N.P. (2020). Enacting Privacy in Internet Standards. University of California, Berkeley.
- ARTICLE 19 (2018). Public interest, private infrastructure: An analysis of the barriers and drivers for adopting human rights standards in the Internet infrastructure industry

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?

Änderungen wurden durch die im Teil ‘Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben’ verursacht.

Verloren Zeit konnte ich an anderen Stellen wieder gewinnen, da die Umstrukturierung unseres Quelltextes es vereinfachte, die Funktionen zum Herunterladen und Bearbeiten von verschiedenen Mailinglisten gleichzeitig zu verbessern.

Ich hatten auch vor, ein interaktives Dashboard für das IETF Normgremium zu erstellen. Jedoch arbeite ich noch daran, da sich die Grösse des Datenvolumen sich als Hindernis dargestellt hat. Da die IETF schon seit 1986 besteht, hat sich eine grosse Menge an Mailverkehr angesammelt die über 20 GB beinhaltet.

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?

Während der Förderphase habe ich mehrere Vorträge im Akademischem Umfeld über das Projekt gehalten. Dadurch ich habe ich viele Kommentare und Rückmeldungen erhalten die sich in meiner weiteren Arbeit an dem Projekt wiederfinden lassen. Ein Beispiel ist, die Mailinglisten Sprachlich zu analysieren und auf Relevanz zu sortieren. Das bedeutet wenn jemand an Geheimhaltung und Privatsphäre oder Klimawechsel interessiert ist, habe ich nun verschiedene Funktionen bereitgestellt um E-Mails zu diesen Themen zu finden.