Richtlinie zum „Software-Sprint“

Exodus 3.0 – Interplanetarische Informationssuche

Schlussbericht

Zuwendungsempfänger:

Zimmermann Zimmermann GbR

DLR PT- Berlin
Eing.am:
30. März 2023
Eingangsnr.: 1261 BK

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS22S32 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)?

Das Internet hat sich als leichte Beute für die Zensur erwiesen: Zugänge werden gesperrt, Server werden vom Netz genommen, DNS-Adressen werden manipuliert. Im vermeintlich offenen Internet findet auch “unsichtbare” Zensur statt. Was von Suchmaschinen nicht als Ergebnis ausgegeben wird, existiert nicht. Eine aktive Blockierung von Inhalten ist dabei gar nicht notwendig. Informationen mögen noch frei verfügbar sein, wenn sie aber nicht recherchiert und aufgefunden werden können, sind sie für niemanden erreichbar. Ohne weitere Anhaltspunkte kann man also einige “zensierte” Inhalte nicht finden und verbreiten, beziehungsweise zugänglich machen. So kommt es vor, dass im Web ganze Internetseiten verschwinden. Sie werden gesperrt oder gar gelöscht. Wir haben in den letzten Jahren mitbekommen, wie leicht ganze Teile des Internets blockiert werden können und wie Suchmaschinen Schlüsselwörter blockieren. Diese Entscheidungen liegen in den Händen einzelner Staaten und Internetgiganten.

Exodus 3.0 tritt diesem Paradigma des Internet entgegen. Unser Ziel war und ist es, einen Beitrag zur Schaffung des Internets der nächsten Generation zu leisten: Eine Suche bei der Datenschutz möglich und Zensur schwierig ist. Mit Exodus 3.0 sollen Bürger*innen sich aus der Zensur befreien und selbst über ihre Sichtbarkeit im Internet verfügen und bestimmen können.

Hauptziel:
Entwicklung eines Tech-Stacks, der interessierten Parteien ermöglicht, ihren eigenen Knotenpunkt zur Indizierung und Bereitstellung von Suchdiensten für verteilte Dateisysteme zu erstellen, mit einem besonderen Fokus auf dem "Interplanetary File System" (IPFS).
Technische Meilensteine:
Wir haben den Tech-Stack erweitert und implementiert, einschließlich einer Reihe von Erweiterungen der re-isearch Engine, um eine transparente Indizierung von "remote" Inhalten über eine Reihe von Protokollen einschließlich IPFS zu ermöglichen. Aufgrund des Datenformats im IPFS wurden auch eine Reihe neuer Dokument-Handler entwickelt und viele bestehende Handler und Klassen wurden modifiziert und erweitert, um den Anforderungen gerecht zu werden. Als Reaktion auf Probleme bezüglich Urheberrechten und potenzieller "rechtlicher Gefahr" wurden auch eine Reihe von Lösungen entwickelt.
Im Laufe des Projekts haben wir uns im Einklang mit dem Ziel, die Lösung der breitesten Öffentlichkeit in der FLOSS-Community zur Verfügung zu stellen, auch darauf konzentriert, den Build und das Deployment durch sowohl erfolgreich verbesserte Builds und Konfigurationen als auch detailliertere Anweisungen und Dokumentationen zu verbessern - in Reaktion auf das Feedback der Benutzer.

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?

Die Zielgruppe sind alle, die über ein gewisses technisches Know-how verfügen und dabei helfen wollen, ein interoperables Suchplattform-Ökosystem zu schaffen. Dabei vertreten mögliche Nutzer die Werte Offenheit, Inklusivität, Transparenz, Privacy, Kooperation und Datenschutz.

Alle Komponenten unserer Lösung sind zu 100% kostenlos und open source (Apache 2.0 Lizenz für die Software und CC 4.0 für die begleitende Dokumentation).

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?

Unsere ursprüngliche Idee war es, einen Prototyp eines Suchknotendienstes für die Suche im IPFS zu erstellen. Während des Projekts entdeckten wir eine Reihe von Problemen.

Der Prozess des Sniffing des DHT Gossip (Auffinden von CID-Hashes) erwies sich als eine von unseren Zielen her schlechte Informationsquelle, die neben interessanten, aber möglicherweise gefährlichen Informationen mit NFTs verschmutzt war.
Wir hatten auch Probleme mit Inhalten, die vermutlich nicht gefunden werden wollen. Da in IPFS so etwas wie eine "robots.txt"- und "not follow"-Direktive fehlt, gab es keine Möglichkeit, uns zu sagen, dass der Inhalt nicht durchsuchbar sein sollte. Dies war eine Reibungsfläche, die über die oben erwähnten Probleme hinausging.
Trotz einiger Experimente fanden wir keine Möglichkeit, Randinhalte zu entdecken. Das Netzwerk erwies sich auch als extrem langsam, und trotz des Ideals der Persistenz erwiesen

sich viele Inhalte als unbeständiger als erwartet - viele CIDs (die Inhalts-URLs) waren ständig unerreichbar.

Wir kamen zu dem Schluss, dass es zwei Möglichkeiten gab:
- Eine Lösung zu entwickeln, die es jedem ermöglicht, IPFS zu indizieren, insbesondere die eigenen Inhalte auf dem eigenen IPFS-Knoten
- Schaffung eines parallelen IPFS-Netzwerks mit expliziten AUPs (Acceptable Use Policies) mit dem formulierten Ziel, durchsucht zu werden: dem EFS.

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 daraus resultierende Software ist frei verfügbar und wird derzeit in ein größeres Moonshot-Projekt (in dem mehrere andere Projekte zusammengeführt werden) integriert, dessen Ziel es ist, eine vollständig verteilte Internetsuche zu schaffen.

Hierfür suchen wir derzeit nach Finanzierungsmöglichkeiten. Dieses größere Projekt hat mehrere Satellitenteams, die sich um München (NONMONOTONIC Networks), Zürich (ETH Student Project House) und Durham North Carolina (Duke University) konzentrieren. Vor diesem Hintergrund reichen wir Anträge sowohl bei der Europäischen Union als auch bei der US-amerikanischen NSF ein und wenden uns außerdem an eine Reihe von gemeinnützigen Stiftungen.

Hat die Arbeit in dem Projekt Dich in Deiner persönlichen, fachlichen Weiterentwicklung unterstützt?

Sehr sogar. Beim IPFS waren wir ein wenig naiv und sahen das Problem als rein technisch an. Obwohl es nie unsere ursprüngliche Absicht war, uns eingehend mit den rechtlichen Fragen zu befassen, machte das Projekt auch uns zu unfreiwilligen internationalen Rechtsexperten.

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?

Unsere ursprüngliche Idee, auch eine öffentliche Suche für das IPFS als Vorzeigeprojekt anzubieten, haben wir fallen gelassen. Wir entwickelten Schnittstellen und Geschäftslogiken, organisierten Serverspace für das Bootstrap (ein großzügiger Grant von AWS), aber selbst unsere Idee des "Offshoring" könnte sich als zu riskant erweisen. Auch unseren Gossip-Stack haben wir aufgegeben, da er mit der Neuausrichtung unseres Ansatzes nicht mehr wirklich relevant war - und eher eine Ablenkung als ein Schritt vorwärts in Richtung unseres großen Ziels, die Werkzeuge für eine verteilte Infrastruktur einer Internetsuche der nächsten Generation zu schaffen, um mit Big Tech zu konkurrieren.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

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

Die Kern-Engine (einschließlich der Support-Plugins für IPFS):
https://github.com/re-lsearch/re-lsearch
Jede Menge Handbücher und Dokumentation unter docs/

Vergleich mit anderen Engines: https://github.com/re-lsearch/re-lsearch/blob/master/docs/re-lsearch-vs-Others.pdf

Ein älterer Vortrag über die Engine selbst: https://archive.fosdem.org/2022/schedule/event/lt_re_lsearch/

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?

Einige Dinge, wie z. B. die Erörterung rechtlicher Fragen und der Arbeitsaufwand zur Vereinfachung von Installationen, wurden zu sehr unterschätzt. Einige Dokumentenverarbeitungsprogramme - wie z. B. ein superschneller ODF-Ingest anstelle von Pandoc - standen ursprünglich nicht auf der Roadmap, sondern entwickelten sich aus der gegenseitigen Absprache mit anderen Projekten.

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 unser Projekt erweitert und ausgebaut. Der Buzz und Hype um LMMs (wie ChatGPT) und die Diskussionen um eine Neuausrichtung der Suche ("Google in Code Red") haben unsere Argumente deutlich verstärkt. Das Timing könnte nicht besser sein!