Das Eigene Webarchiv

Schlussbericht

Zuwendungsempfänger:

Gerben van den Broeke

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS18S20 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Wir sind von einem Web abhängig, das instabil und vergesslich ist. Jede Website wird von einem einzigen Provider bereitgestellt und wenn dieser aufhört, den Zugang anzubieten, oder die Inhalte ersetzt, hat niemand eine Kopie der Seite, wie sie zuvor war. Noch problematischer ist, dass man dann sogar nicht einmal zu einer alten Version verlinken oder Änderungen erkennen kann.

Das eigene Webarchiv ermöglicht jedem Seite zu archivieren, wieder zu veröffentlichen und teilen; und immer ohne abhängen zu müssen von z.B. ein Cloud Service.

Hauptteil dieses Projekts war ein Browser-Erweiterung zu machen, mit welchem man Schnappschüssen von Webseite machen kann. Um diese Schnappschüsse irgendwo zu archivieren, wieder zu Nützen, und möglicherweise mit anderen zu teilen, ist auch ein Serverteil entwickelt.

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

Die Zielgruppe sind alle, die das Internet als Informationsquelle nutzen wollen und für längere Zeit Zugang zu diesen Informationen haben möchten. Besonders wichtig ist es für Leute mit Interesse an Dokumenten bei denen ein Risiko besteht, dass sie zensiert, zurückgenommen, oder manipuliert werden. Genau dieses Aspekt gibt das Projekt eine hohe gesellschaftliche Relevanz.

Ausführliche Darstellung der Ergebnisse

In die sechs Monate habe ich das Algorithmus für das Speichern der Webseiten, *freeze-dry* genannt, von neuem entwickelt, mit genaue Beachtung für neue und alte Webstandards um so viel möglich Webseiten adäquat bewahren zu können. Die *WebMemex* Browsererweiterung nutzt jetzt dieses neue Algorithmus, und ist selbst auch gründlich verbessert.

Das Serverteil ist entwickelt als Nextcloud-App, und macht auf die Nextcloudserver gespeicherte Schnappschüsse verfügbar über das Memento-Protokoll, genau so wie existierende Webarchiven (z.B. das Internet Archive).

Ein nicht erreichte Meilensteine ist der Funktionalität den Unterschied zwischen der neuen und alten Version einer Seite hervorzuheben. Es ist aber in die Erweiterungsinterface sehr einfach gemacht um zwischen die Versionen zu wechseln.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Nutzer der Browsererweiterung können jetzt einfach Webseite speichern, durchsuchen, und den Schnappschuss als HTML-Datei speichern und teilen. Für den Open-Source-Community ist es nützlich dass das *freeze-dry* Algorithmus als unabhängiges Modul publiziert ist, denn andere Projekten es nutzen können.

Das Projekt kann natürlich immer weiterentwickelt werden. Einerseits kann die gerade entwickelte Software verbessert werden, und andererseits kann man neue Anwendungen finden mit die gespeicherte Seiten und persönliche Webarchiven, zum Beispiel Ordnungs- oder Notizen-Systemen. Selbst Arbeit ich unbedingt weiter in diesen beide Richtungen.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Ein logischer Plan wäre gewesen, die Browsererweiterung und Serverteil direkter miteinander zu verknüpfen. Dieser Plan hätte zwei teilen: einerseits die Möglichkeit zu geben die Schnappschüsse direkt auf den Server zu speichern, statt in die Browsererweiterung; und andererseits die Browsererweiterung nicht nur durch ihre lokale Schnappschüsse suchen zu lassen, aber auch durch deinen Archivserver, und selbst durch den geteilten Archiven von anderen.

Ich habe mit beiden Möglichkeiten experimentiert, aber es würde mehr Arbeit brauchen um das kohärent funktionieren zu lassen. Um die Nutzer nicht zu verwirren mit halbherzige Funktionen, hab ich diese Integration vorerst aufgegeben. Die Nutzer können immer noch mittels drag&drop ihre Schnappschüsse auf ihren Nextcloud-Server hochladen, und können mit existierende Memento-Clients ihre oder anderleuts Archivservers nutzen.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Allgemeine Projektinfo: https://blog.webmemex.org/

Quellcode der Browsererweiterung:
https://github.com/WebMemex/webmemex-extension

Code und Erklärung der Technologie entwickelt für das speichern der Seiten:
https://github.com/WebMemex/freeze-dry

Code und Erklärung des Nextcloud-Apps um ein eigene Archiv-Server zu machen: https://code.treora.com/gerben/nextcloud-memento

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Das entwicklen des *freeze-dry* Algorithmus hat mehr Zeit gekostet als geplant, aber außerdem scheint alles ungefähr im erwarteten Bereich geblieben zu sein.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Es gab Interesse von Leute bei Mozilla um die Hauptfunktionaliteit, das speichern von Webseiten mittels freeze-dry, direkt in Firefox selbst einzubauen (diese Funktion gibt es in Theorie gerade, aber ist kaum nutzbar). Dafür war ich eingeladen beim ihre All Hands meeting in San Francisco, um über diese Möglichkeit zu reden. Dass eine Integration wie dieses viel besser sein würde als das Publizieren einer Browsererweiterung, war einer der Gründe meinen Fokus mehr in der Richtung von das erstellen der Speicherungstechnologie als in der Richtung von die Erweiterung um es herum zu verlegen. Obwohl diese mögliche Integration noch nicht weitergemacht ist, denke ich dass diese Fokusänderung ein gute Wahl war, weil gute Softwarekomponente sehr wertvol sein für das Open-Source-Ökosystem.