Richtlinie zum „Software-Sprint“

City Census - Automatisierter Open Data City Census

Schlussbericht

Zuwendungsempfänger:

Mila Frerichs

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS17S22 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Mit meinem Projekt wollte ich einen ersten Ansatz in der Bewertung von Open Data Portalen und Open Data Datensätzen in Deutschland erstellen. Darüber hinaus wollte ich eine Diskussion über Metadaten und die Pflege von Open Data Portalen herbeiführen.

Ich wollte Städte motivieren mehr und bessere Daten zu veröffentlichen und über die Qualität der vorhanden Daten nachzudenken. Ein Ranking ist immer auch ein Vergleich und hat immer Gewinner und Verlierer. Ich wollte die sowohl Gewinner als auch Verlierer motivieren besser zu sein und sich gegenseitig etwas abzuschauen.

Außerdem wollte ich Transparenz herzustellen in einem Bereich, der üblicherweise nur einem sehr beschränkten Nutzerkreis vorbehalten ist. Die verwendete Technik erlaubt nicht so ohne weiteres jedem Bürger einen Einblick in die Qualität der Portale.

Meilensteine:

- Auslesen der Quellen
- Bewertung der Metadaten
- Was kann als Scoring herangezogen werden
- Erster Vorschlag für ein Scoring
- Feedback zum ersten Scoring

- Treffen mit Städten zum Scoring

- Govdata.de oder offenedaten.de als Quelle

- Mehr Städte auf offendedaten.de

- Automatisierung

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 dieses Projektes sind vor allem die Mitarbeiter von Städten und Kommunen die Open Data Programme betreiben. Darüber hinaus ist der „normale“ Bürger Zielgruppe für den Stand seiner eigenen Stadt im Ranking.

Die Portale der Städte zu bewerten ist ein Versuch mit diesen ins Gespräch zu kommen und bessere Daten bereitzustellen. Dieses Ziel wurde mehr als erreicht, da viele Städte über das Projekt Bescheid wissen und die Ergebnisse in Ihre Arbeit einfließen lassen.

Darüber hinaus hat das Projekt eine Diskussion innerhalb der Open Data Gemeinde ausgelöst zur Qualität von Datensätzen und der Pflege der Metadaten von Open Data Portalen.

Das Projekt hat einen direkten Bezug zu Civic Tech, da die Nutzung von Open Data ein essentieller Bestandteil von Civic Tech ist.

Ausführliche Darstellung der Ergebnisse

Der erste Prototyp ist fertig und wurde vielfach aufgerufen. Er hat, wie erhofft eine Diskussion innerhalb der Städte, die Open Data Programm haben, ausgelöst. Ich befinde mich in Diskussion mit mehreren Kommunen über die weitere Arbeit an dem Projekt und die Verbesserung der Metadaten der Portale.

Es werden täglich die Open Data Portale Deutschlands abgefragt und nach einem eigens entwickelten Scoring gerankt. Das Projekt läuft autark weiter und benötigt keine Eingriffe meinerseits.

Das war ein großes Ziel des Projektes, dass ohne Mehraufwand für die Kommunen und keine großen Pflegeaufwand auf meiner Seite entsteht.

Mit Hilfe der Cloud und AWS Lambda und S3 kann kostengünstig, schnell und effektiv ein Ranking jeden Tag erstellt werden.

Es wurden sogenannte Harvester geschrieben die alle Open Data Portale auslesen und die Metadaten auf offenedaten.de abspeichern. Von dort können sie einfach ausgelesen werden.

Die Automatisierung ist das Herzstück des Projektes und hat die meiste Energie und Arbeit benötigt. Die Bewertung von 20.000 Datensätzen und 146 Städten ist eine große Aufgabe und konnte mit Hilfe der Cloud gelöst werden.

Alle Daten werden auf Github veröffentlicht und können mit Hilfe des Versionssoftware git nachverfolgt werden. Es werden darüber hinaus die Plätze des Rankings gespeichert um einen Verlauf sehen zu können.

In mehreren Iterationen ist das Scoring entstanden. Bei der Erstellung des ersten Scorings sind Ergebnisse entstanden, die nicht einfach zu erklären waren.

Durch mehrere Analysen der vorhandenen Datengrundlage könnte das scoring nach und nach verbessert werden. Ingesamt gab es mindestens 5 Versionen für das Scoring.

Das Scoring ist ein erster Schritt zur Messung der Qualität deutscher Open Data Programme und deren Umsetzung. Es kann und soll weiter verbessert werden in Zusammenarbeit mit Städten und Kommunen. Einige Städte haben sich bereits dazu bereit erklärt.

Zusätzlichen Erkenntnisse

Die Workshops, organisiert von der Open Knowledge Foundation, begleitend zum Projekt waren sehr hilfreich und haben einen bleibenden Einfluss auf meine weitere Arbeit. Darüber hinaus waren sie wertvoll in der letzten Umsetzungsphase meines Projektes.

Leider konnten nicht alle Meilensteine erreicht werden, da das Feedback während der Förderphase nicht meinen Hoffnungen entsprach. Ein wichtiger Meilenstein war eine Diskussion während der Projektphase anzustoßen und mich mit mehreren Städten zu treffen. Leider habe ich kaum bis gar kein Feedback bekommen, so dass ich mich auf die Erstellung des Prototypen konzentriert habe. Das Feedback kam aber sehr schnell nach der ersten Vorstellung der Ergebnisse beim Demo-Day und in der Vorbereitung auf diesen, als ich mit dem fertigen Projekt auf die Top 5 zugegangen bin.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Zielgruppen waren die Open Data Beauftragten der Städte und Kommunen in Deutschland und interessierte Bürgerinnen und Bürger.

Die Zielgruppe der Städte hat ein erstes Benchmark bekommen. Hier ein paar Zitate von Mitarbeitern der Open Data Programme deutscher Städte:

„Es kann hilfreich sein sich selbst als Open-Data-Kommune zu verbessern und man bekommt gegebenenfalls Hinweise auf Fehlerpotenziale sowie Anregungen für Verbesserungen.„

„Das Ranking spiegelt nämlich aus unserer Sicht die wahren Leistungen und Qualitäten der Anbieter offener Daten wieder.“

„Das ist eine schöne Motivation für uns sich für die Bereitstellung „Offener Daten" in der Verwaltung der Hansestadt Rostock so intensiv zu engagieren.“

Es ist eine Diskussion angestoßen worden, wie Daten besser bereitgestellt werden können in den Portalen und das sich die Städte untereinander besser abstimmen könnten.

Es hat sich eine Gruppe von Städten formiert, die an diesem Problem arbeiten wollen und durch den Prototypen sich in Bewegung gesetzt haben.

Das Projekt soll und wird weiterentwickelt. Es gibt eine Roadmap wie das Projekt weitergehen kann und es wird gerade nach Fördermöglichkeiten gesucht.

Aus der Open Source-Stellung ergibt sich die Möglichkeit sehr genau zu schauen, wie das Scoring entstanden ist und es können Vorschläge gemacht werden, wie dieses verbessert werden kann. Durch das bereitstellen der Software als Open Source wird eine erweiterte Transparenz erreicht von der alle beteiligten Profitieren. Auch eine weitergehende Entwicklung des Projektes wird durch die Bereitstellung als Open Source vereinfacht und gefördert.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Die Organisation von Workshops mit Städten und Kommunen war ein wichtiger Meilenstein der leider nicht erreicht werden konnte. Das lag zum einen an dem nicht erfolgreichen ersten Feedback von Städten und Kommunen. Nach mehreren Versuchen per Email und auf Konferenzen Feedback und Gespräche zu führen, musste dieser Teil leider auf Eis gelegt werden. Es wurde klar, dass ein sehr früher Prototyp oder auch nur die Diskussion zu diesem Thema nicht sehr viel Zuspruch von Seiten der Verwaltungen erreichen konnte. Deswegen habe ich meine Zeit und Kraft darauf verwendet einen lauffähigen Prototypen zu entwicklen, der auch nach der Förderphase funktionieren wird und weiterhin für Diskussionen sorgt.

Um Diskussionen zu provozieren habe ich einige Kriterien mit in das Scoring aufgenommen und überspitzt um eine Reaktion bei den Kommunen und der Community hervorzurufen. Dies ist mir in der Rückschau gelungen. Durch die Überspitzung habe ich Reaktionen hervorgerufen, die zu Diskussionen führten. Nach der Erklärung wie das zustande kam auch sehr konstruktive Diskussionen.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Der Programmcode ist auf GitHub veröffentlicht inklusive aller Daten die täglich generiert werden. Es gibt eine Webseite mit der Darstellung der Ergebnisse und weiteren Informationen zum Projekt und dem Scoring. Diese wird ständig mit mehr Informationen erweitert. Darüber hinaus befindet sich die komplette Dokumentation auf GitHub.

Beim Open Data Day in München wurde das Projekt im Rahmen der Keynote vorgestellt, die Präsentation und das Video ist ebenfalls auf GitHub veröffentlicht.

Ich werde im Mai in Detroit einen Talk über das Projekt und die Herangehensweise im Rahmen der Konferenz „Good Tech Fest“ halten.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Der Kosten - und Arbeitspläne wurde eingehalten. Es gab keine Ereignisse die einen Mehr- oder Minderaufwand erzeugt haben.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Es gab während der Projektphase ein neues Projekt wie man Open Data allgemein bewerten kann. Die Open Data Measurement Group hat sich grundsätzlich mit dem Thema beschäftigt

Ich hatte ein Telefongespräch mit eine, der Initiatorin und wir haben grundsätzlich über das Thema und mein Projekt gesprochen.

Da die Diskussion gerade angefangen hatte und sich mehr auf die Bewertung von Ländern bezog hatte es keine Einfluss auf das Projekt.

Wir haben auch über meinem Ansatz zum Scoring gesprochen und es wurde grundsätzlich als gut betrachtet.