Open Legal Data

Schlussbericht

Zuwendungsempfänger:

Open Legal Data - Ostendorff Blume Langhabel GbR

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS18S40 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)?

Open Legal Data ist eine freie Plattform, die juristische Inhalte der Öffentlichkeit nachhaltig so zur Verfügung stellt, dass sie leicht sowohl von Personen als auch von Maschinen genutzt werden können. Dazu bedienen wir uns neuesten Verfahren der automatischen Aufbereitung und Erschließung von öffentlichen Urteilen und Rechtsnormen. Somit werden nicht nur die Dokumente zugänglich gemacht, sondern auch die darin enthaltenen Information. Dadurch soll der Rechtsstaat transparenter werden und der Umgang mit Rechtsthemen für Nicht-Juristen erleichtert werden, wobei wir zunächst technisch versierte Bürger ansprechen. Das Projekt verpflichtet sich dabei den Prinzipien der Open Data Bewegung.

Ein Urteil wird im Namen des Volkes gesprochen, ist damit Ausdruck des Demokratieprinzips (Art. 20 GG) und nicht urheberrechtlich geschützt. Dennoch sind viele Urteile nur unzureichend für die Allgemeinheit zugänglich. Dies führt dazu, dass die Akzeptanz der Justiz abnimmt und Unverständnis für verschiedene Urteile herrscht. Neben der Unzugänglichkeit fehlt eine überzeugende und transparente amtliche Statistik der Rechtsprechung. Dies wird u.a. auch beim Deutschen Juristentag thematisiert und kritisiert. Mit unserem Projekt machen wir die Rechtsprechung für alle zugänglich und mit Hilfe von unabhängigen und auf Open Data basierten Statistiken die Rechtsprechung transparenter.

Der Kern des Projekts ist eine Website inkl. API. Über die Website werden alle Inhalte optisch ansprechend und übersichtlich dargestellt. Zusätzlich ermöglicht eine Suchfunktion das Finden von relevanten Inhalten. Neben der Website wird der Austausch von juristischen Daten vor allem durch eine API ermöglicht. Über die API sind nicht nur alle Inhalte in maschinenlesbarer Form abrufbar, sodass sie von anderen Diensten genutzt werden können, sondern wir bieten auch eine Möglichkeit, Inhalte mit Zusatzinformationen, wie Erklärungen und Streitwerte, anzureichern (Open Annotations). Das Projekt ist in Python geschrieben mit dem Django Framework einschließlich REST-API Addon als Grundgerüst. Elasticsearch wird für die Suchfunktion benutzt. Als Datenbank ist MySQL im Einsatz. Des Weiteren stellen wir API Clients und Jupyter Notebooks mit direktem Zugriff auf alle Daten zur Verfügung, um weiterführende Analysen zu erleichtern.

Für die Umsetzung des Projektes haben wir uns fünf Meilensteine gesetzt:

1) Open Source Veröffentlichung mit angemessener Codequalität sowie integrierten Test und grundlegender Dokumentation

2) Vollständige Verfügbarkeit von deutschen und europäischen Rechtsnormen und Verordnungen

3) Über 100.000 Urteilstexte inklusive Metadaten von deutschen und europäischen Gerichtshöfen (darunter EuGH, BGH, BVerfG, BVerwG) sind maschinenlesbar über die API abrufbar

4) Eine öffentlich konfigurierbare Data Processing Pipeline, die es externen Entwicklern ermöglicht gezielt Informationen aus den Daten zu extrahieren, z.B. Streitwert oder verwendete Rechtsnormen

5) Veröffentlichung eigener Analysen

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 primäre Zielgruppe von Open Legal Data sind interessierte Aktivisten, Journalisten und Entwickler, die mit unserer Hilfe Analysen, Artikel, Apps etc. erstellen können. Um unsere Zielgruppen zu erreichen, haben wir u.a. Open Legal Data beim 35. Chaos Communication Congress[1] vorgestellt. Mit der Veröffentlichung eines Datensatz von Gerichtsurteilen haben wir den Global Legal Tech Hackathon[2] unterstützt. Mit der Hilfe unserer Daten gewann das Team VizLaw den Legal Tech Hackathon in Berlin[3]. Auf diese Weise konnten wir den Nutzen von Open Source und Open Data für den juristischen Bereich demonstrieren und nachhaltig eine Diskussion auch über die Grenzen der „Open Data-Szene“ hinaus anstoßen.

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?

Durch die Förderung konnten wir aus einer Idee im Laufe des Projektzeitraums einen nutzbaren Prototypen entwickeln. Nahe zu alle selbstgesteckten Meilensteine konnten erreicht werden. Der gesamte Quellcode der Open Legal Data Plattform ist auf GitHub unter der freien MIT Lizenz veröffentlicht[4] (MS1). Alle dazugehörigen Module, wie Zitaterkennung oder Named Entity Recognition, werden ebenfalls zur Verfügung gestellt. Über die Plattform können nicht nur Inhalte abgerufen werden. Unser Annotation API (Data Processing Pipeline) ermöglicht es Dritten, die auf der Plattform enthalten Dokumente mit zusätzlichen Informationen anzureichern (MS4). Wie diese

[1] #35C3 Open Legal Data - Lighting Talk, https://www.youtube.com/watch?v=5KHcXCmqfiE
[2] Global Legal Hackathon 2019, https://globallegalhackathon.com/
[3] VizLaw uses our API and wins Berlin Legal Tech Hackathon 2019, http://openlegaldata.io/events/2019/02/27/vizlaw-wins-berlin-legal-tech-hackathon.html
[4] GitHub – Open Legal Data Plattform, https://github.com/openlegaldata/oldp

Funktion genutzt werden kann, demonstrieren wir beispielhaft mit Streitwerten[5]. Auch das im Rahmen des Software Sprints geschaffene Named Entity Recognition Modul spielt als Schritt der Data Processing Pipeline Daten über die API in die Datenbank ein. Das inhaltliche Angebot konnten wir auf mehr als 250.000 juristische Dokumente ausbauen. Während des Projektzeitraums haben wir uns inhaltlich auf Urteilstexte fokussiert, sodass MS3 mit mehr als 150.000 Urteilen übertroffen werden konnte. Bezüglich Gesetzestexte, konnte MS2 nicht vollständig erreicht werden, da die Gesetze der Länder bisher noch nicht über die Plattform abrufbar sind. Auf Basis der gesammelten Inhalte konnten wir bereits eigene Analysen, wie z.B. ein Zitatnetzwerk deutscher Gerichtsentscheidungen[6], erstellen (MS5).

Insgesamt hat die Förderung uns ermöglicht große Fortschritte mit unserem Projekt zu machen. Insbesondere auch konnte die Kernwebsite von einem Ecosystem an Komponten, Beispielen und Anwendungen umgeben werden. Durch die gemeinsamen Workshops und Coaching konnten wir uns professionalisieren und ein Netzwerk aufbauen, sodass wir auch langfristig das Projekt vorantreiben können.

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

Als eines der ersten Open Source Projekte im juristischen Bereich konnten wir nachhaltig für die Vorteile der Open Source Idee werben. Das Beispiel VizLaw zeigt wie Innovation durch Open Data gefördert werden kann. Um Open Legal Data auch langfristig betreiben zu können, sind wir bereits in Gesprächen, um das Projekt auch auf andere Länder zu erweitern. Über ein universitäres Forschungsprojekt soll zusätzlichen der wissenschaftliche Nutzen nachhaltig erhalten bleiben.

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?

Es wurden keine geplanten Arbeiten gestoppt. Während des Förderzeitraums wurden lediglich einige Teilaufgaben anders priorisiert. Zum Beispiel haben wir uns auf ausschließlich Urteile als Inhalte fokusiert. Andere Dokumentarten, wie Gesetze, werden wir jedoch in Zukunft auch hinzufügen.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

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

Die erste Anlaufstelle für mögliche Nutzer ist unsere Website, auf welcher alle Inhalte abgerufen werden können und wir über unsere Arbeit in einem Blog berichten. Neben der Website sind wir u.a.

---

[5] Litigation Value Extraction, https://github.com/openlegaldata/oldp-notebooks/blob/master/notebooks/litigation_value.ipynb

[6] Analysis: Citation Network of German Court Decisions, http://openlegaldata.io/de/citation-network.html

auf Twitter[7] präsent, dort konnten wir unsere Followerzahl innerhalb des Förderzeitraums um 620% steigern. Für die Entwickler-Community sind alle unsere Arbeiten auf GitHub[8] verfügbar.

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?

Alle Teilaufgaben wurden wie geplant bearbeitet.

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?

Die Veröffentlichung des Projektes OffeneGesetze.de im Dezember 2018 durch die Open Knowledge Foundation veranlasste uns dazu, den Schwerpunkt von Open Legal Data auf insbesondere Rechtsprechung zu verschieben.

[7] https://twitter.com/openlegaldata/
[8] https://github.com/openlegaldata/