

In der Staatsbibliothek zu Berlin der Stiftung Preußischer Kulturbesitz sind zum nächstmöglichen Zeitpunkt drei Stellen einer/eines

**Software-Entwicklerin/Software-Entwicklers
mit Schwerpunkt Machine Learning/Information Retrieval**

**im Projekt "QURATOR. Curation Technologies"
bis Entgeltgruppe E13 TVöD
- je nach Qualifikation -
Kennziffer: SBB-IDM-7-2018**

befristet bis zum 30.09.2021 zu besetzen.

Je nach Qualifikation und Berufserfahrung kann eine übertarifliche Stufenzuordnung und/oder eine IT-Fachkräftezulage gewährt werden.

Die Staatsbibliothek zu Berlin (SBB) ist eine wissenschaftliche Universalbibliothek, die zu den größten deutschen Bibliotheken zählt. Gemeinsam mit Partnern aus Wissenschaft und Wirtschaft führt sie das vom BMBF geförderte Projekt „QURATOR. Curation Technologies“ durch.

Das Projektteam an der SBB stellt ein dynamisches und mit modernen Technologien arbeitendes Entwicklungsumfeld dar, welches an ein eigenes Rechenzentrum mit einer leistungsfähigen Virtualisierungslandschaft angebunden ist. Aktuell werden ca. 2,5 PB an Daten, darunter Digitalisate und durch OCR erzeugte Volltexte, bereitgehalten. Wir suchen innovative Mitarbeiter/innen, die über Erfahrungen und Kenntnisse in der Softwareentwicklung auf Basis von Machine-Learning-Technologien verfügen und mit einem hohen Anteil an Eigeninitiative die verschiedenen Aufgaben wahrnehmen sowie an der technologischen Konzeption und dem Softwaredesign teilnehmen.

Aufgabengebiet:

Sie bearbeiten im Team das Teilgebiet „Automatisierte Kuratierungstechnologien für das digitalisierte kulturelle Erbe“ in welchem die Konzeption, Implementierung und Evaluierung von Softwarekomponenten zur semantischen Analyse, Inhaltsanreicherung und Profilierung der digitalisierten Bestände der Staatsbibliothek zu Berlin umgesetzt werden soll. Hierbei sind die drei Stellen in den Softwareentwicklungsbereich der Staatsbibliothek eingegliedert, durch welchen sie auch unterstützt werden.

Im Rahmen des Projektes sind folgende, beispielhafte Bereiche zu bearbeiten, die sich auf die drei Stellen verteilen:

- Dokumentenanalysen z.B. mittels Methoden des Clusterings
- Automatisierte Korrektur von Volltexten aus der OCR mittels Heuristiken, n-Grammen oder Anagram Hashing
- Multimodale Struktur- und Inhaltserkennung mithilfe von Verfahren des maschinellen Lernens (z.B. CNN)
- Entwicklung von Verfahren zur Toponym-Auflösung, zur Disambiguierung bzw. Verlinkung verschiedener Entitäten
- Anreicherung bestehender Metadatenbestände um semantische Metadaten (Inhaltsbeschreibungen, Named Entitys, Word Embeddings etc.)
- Erstellung entsprechender Testdatensets und Einbeziehung dieser zur Qualitätssicherung

Anforderungen:

Abgeschlossenes wissenschaftliches Hochschulstudium der Informatik, Informationswissenschaften, Computerlinguistik, Data Science, Data Analytics, Angewandten Statistik, Angewandten Mathematik oder einer vergleichbaren technisch-naturwissenschaftlichen bzw. mathematischen Fachrichtung mit IT-Ausrichtung oder auf die Aufgaben bezogene gleichwertige Fähigkeiten und Erfahrungen • sehr gute Kenntnisse in der Programmiersprache Python und der einschlägigen Data-Science-Toolkits (z.B. numpy, pandas, scikits etc.) • gute Kenntnisse in Frameworks für maschinelles Lernen (z.B. Keras, TensorFlow etc.) • Erfahrungen in der Erstellung von technischen Konzepten • sehr gute Auffassungsgabe sowie selbstständiges, systematisches, lösungs- und ergebnisorientiertes Arbeiten sowie die Fähigkeit sich in fachwissenschaftliche Fragestellungen einzuarbeiten • gute Team- und Kommunikationsfähigkeit, hohe Kundenorientierung und Verantwortungsbewusstsein

Erwünscht:

Kenntnisse verschiedener Datenbankparadigmen (SQL, XML, Key-Value, Graphen...) und Information-Retrieval-Systeme (insb. Solr) • gute Kenntnisse im Umfeld von Linked-Open-Data-Technologien (z.B. SPARQL, RDF) • Kenntnisse von Datenmanagement- und Mininigwerkzeugen (z.B. OpenRefine, catmandu, Trifacta, SAS Enterprise Miner etc.) • fundierte Kenntnisse von aktuellen Softwarearchitekturen und Programmierparadigmen wie z.B. Microservices und REST/RESTful APIs • Kenntnisse im Linux-Umfeld • Kenntnisse gängiger Formate und Protokolle in Gedächtniseinrichtungen (z.B. METS/MODS, ALTO, TEI, OAI-PMH, SRU etc.) • sicherer Umgang mit gängigen Entwicklungstools (IDEs, Versionsverwaltungssoftware und Build-Management-Tools) • gute Deutschkenntnisse

Wir bieten:

Flexibles Arbeitszeitmodell auf Basis einer 39-Stunden-Woche sowie ansprechende Weiterbildungsmöglichkeiten, um Sie sowohl fachlich als auch persönlich zu fördern

Die Stiftung Preußischer Kulturbesitz begrüßt Bewerbungen von Menschen aller Nationalitäten. Als interdisziplinäre Kultur- und Forschungseinrichtung bietet die SPK familienfreundliche Arbeitsbedingungen und gewährleistet die Gleichstellung von Frauen und Männern. Schwerbehinderte Menschen werden bei gleicher Eignung besonders berücksichtigt.

Eine Besetzung mit Teilzeitbeschäftigten ist grundsätzlich möglich.

Bitte richten Sie Ihr Anschreiben an die Generaldirektorin der Staatsbibliothek zu Berlin, Barbara Schneider-Kempf, und senden dieses zusammen mit aussagekräftigen Bewerbungsunterlagen ausschließlich in schriftlicher Form (keine E-Mail und keine elektronischen Datenträger) unter Angabe der Kennziffer **SBB-IDM-7-2018** bis zum **29. November 2018** an:

Stiftung Preußischer Kulturbesitz, Personalabteilung, Sachgebiet I 2c, Von-der-Heydt-Str. 16-18, 10785 Berlin

Ansprechpartner für Fragen zum Aufgabengebiet: Herr Dr. Zellhöfer (Tel.: 030 - 266 43 2200),

Ansprechpartnerin für Fragen zum Bewerbungsverfahren: Frau Mohnke (Tel.: 030 - 266 41 1700).