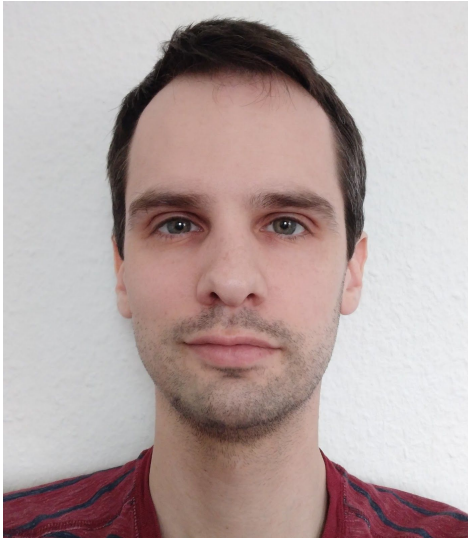# Aaron Christianson

Job title: Software developer

Organisation: University Library Johann Christian Senckenberg

Twitter: @ninjaaron_0

Abstract:

Title: Back to the Source: Recovering Original (Hebrew) Script from Transcribed Metadata

Due to technical constrains of the past, metadata in languages written with non-Latin scripts have frequently been entered using various systems of transcription. While this transcription is essential for data curators who may not be familiar with the source script, it is often an encumbrance for researchers in discovery and retrieval, especially with more complex forms of transcription, as are common with Arabic and Hebrew scripts. The University Library Johann Christian Senckenberg has a very large Judaica collection with many works in Hebrew and Yiddish. Until 2011, all these works were catalogued with transcription only. We are currently developing an open-source system to aid in the automatic conversion of Hebrew transcription back into Hebrew script using a multifaceted approach. Because of inherent ambiguities in the transcription, the initial step is generating a list of theoretically possible original readings and using various methods to narrow down the results. This includes simple spell-checking, matching against extent Hebrew metadata, finding sources where transcription is available alongside the original script, and eventually culminating in the application of phonological analysis. With the generated corpus, we will use machine learning techniques to train a model which can convert arbitrary texts with a high degree of accuracy. All of our tools are published as open source. During the presentation, we will

discuss the progress we've made so far, the challenges we've faced, and present an outlook on the next steps. In addition, we present other possible applications of the technology.

Bio:
Originally from Wisconsin, Aaron taught Biblical Hebrew for several years following the completion of his MA in Hebrew Bible and Ancient Near East at the Hebrew University of Jerusalem, but turned his long-time programming hobby into a career when he was hired as a software developer by the University Library Johann Christian Senckenberg in Frankfurt in 2016 to help solve special challenges in the retrieval of resources in non-Latin scripts (Hebrew and Yiddish specifically). Aaron loves Python, open source software, Semitic languages and fancy computer keyboards.