

TEIHencer - Enhance your TEI-Documents

Andorfer, Peter

peter.andorfer@oeaw.ac.at
ACDH, Österreich

Karner, Stefan

stefan.karner@onb.ac.at
Österreichische Nationalbibliothek

Die Forschungstätigkeiten Georeferencing und Entity Linking sind wichtiger Bestandteil vieler DH-Projekte. Webservices/APIs und Tools versuchen diese Tätigkeiten zu vereinfachen und zu beschleunigen. Eines der bekannteren Tools, wenigstens im deutschsprachigen Raum ist dabei vermutlich der ‘DARIAH-DE Datasheet Editor’¹. Dieser zeichnet sich durch seine einfache Benutzung aus, sei es was den Datenimport (ausfüllen einer Tabelle oder Hochladen einer CSV-Tabelle) betrifft oder die anschließende Disambiguierung/Verifizierung der vom ‘Getty Thesaurus of Geographic Names’² zurückgelieferten Treffer über ein Graphical User Interface.

Das Projekt TEIHencer greift diese Vorzüge des ‘DARIAH-DE Datasheet Editors’ auf und versucht diese einerseits mit der ‘TEI-Welt’ zu verknüpfen sowie mit GeoNames und der GND zwei alternative Normdaten Ressourcen einzubinden.

Konkret handelt es sich bei TEIHencer³ um ein Plug-In zu dem Python/Django basierten prosopographisch-geographischen Informationssystem APIS⁴. Mit Hilfe des TEIHencers ist es möglich, XML/TEI kodierte Texte in denen Lokalitäten, Orte ausgezeichnet sind, über eine Webformular in APIS zu importieren. Während des Imports werden die Orts-Entitäten entsprechend eines vom Benutzer wählbaren X-Path Ausdrucks geparkt, gegen GeoNames und GND abgeglichen und im Falle von Übereinstimmung angereichert und in einer relationalen Datenbank gespeichert. Die gespeicherten Entitäten können anschließend über das APIS-Web-Interface im Falle mehrerer Treffer disambiguiert werden. Dies erfolgt über eine Kartendarstellung, in welcher die verschiedenen Treffer zu einer Entität aufscheinen. Darüber hinaus können über das APIS-Web-Interface noch weitere Informationen zu den Entitäten ergänzt (z.B. alternative Schreibweisen, Datierungen) sowie die einzelnen Entitäten miteinander in typisierte Beziehungen gesetzt werden (z.B. Ort A ist Nachfolger von Ort B.; Ort A ist Teil von Ort B).

Die mit Hilfe von TEIHencer angereicherten Daten können dann wieder als XML/TEI Dokument (kodiert als <tei:listPlace> Element) exportiert bzw. über HTTP GET request abgerufen und so etwa in andere Applikationen eingebunden werden.

Im Zuge der Posterpräsentation soll der TEIHencer der einschlägigen DH-Community vorgestellt werden und zwar an dem konkreten Fallbeispiel der ‘Andreas Okopenko: Tagebücher aus dem Nachlass (Hybridedition)’⁵. Dabei handelt es sich um ein digitales Editionsprojekt, das eine Auswahl der Tagebücher Andreas Okopenkos im Zeitraum von 1949 bis 1955 inhaltlich erschließen und einem breiteren Publikum zugänglich machen möchte. Einer der Schwerpunkte des Projekts liegt hierbei auf der inhaltlichen Erschließung des örtlichen Wirkungs- und Schaffensraums des Nachkriegsavantgardisten, indem nicht nur erwähnte Orte (<tei:placeName>), sondern nach Maßgabe auch Werke und Organisationen (<tei:title> und <tei:orgName>) mit geographischen Normdaten verknüpft werden, um so ein umfassenderes Bild von Okopenkos kulturellem Kontext vermitteln zu können.

Neben der eigentlich Applikation und des konkreten Use-Cases wird am Poster auch das Konferenzthema ‘Kritik der Digitalen Vernunft’ bzw. das Subthema ‘Kritik digitaler Angebote, Projekte und Werkzeuge’ in Form der Frage nach der Nachhaltigkeit des vorgestellten Tools reflektiert. Eine solche glauben wir nämlich insofern gewährleisten zu können, als das Tool a) in ein konkretes Projekt (Okopenko) eingebettet ist, b) einen weit verbreiteten Standard (TEI) unterstützt, c) auf bestehende Eigenentwicklungen (APIS) aufbaut und d) Teile des Codes als selbstständige Module (TEI-Modul als python-package) konzipiert sind, die auch jenseits der konkreten Applikation Anwendung finden können. Darüber hinaus, e) ist der gesamte Code auf GitHub publiziert [3].

Fußnoten

1. <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>
2. <https://geobrowser.de.dariah.eu/edit/index.html>
3. <https://github.com/acdh-oeaw/teihencer>
4. <https://github.com/acdh-oeaw/apis-core>
5. <https://www.onb.ac.at/bibliothek/sammlungen/literatur/forschung/projekte/andreas-okopenko-tagebuecher-aus-dem-nachlass-hybridedition/>