

Revrit Verfahren zur automatisierten Retransliteraton von Hebraica-Metadaten

ZIEL

Verbesserung des Nachweises und der Auffindbarkeit von Hebraica

Ausgangslage:

- Komplizierte, für Nutzer*innen häufig unverständliche Umschriftregeln, basierend auf verschiedenen Standards, was die Auffindbarkeit von Hebraica in elektronischen Katalogen erheblich erschwert.
- Katalogisierung von Hebraica mit Originalschrift erst seit etwa 2010 üblich, aber weiterhin nicht an allen Bibliotheken möglich

WIE?

- Automatisiertes Verfahren zur Anreicherung von Titeldaten mit hebräischer Originalschrift
- Bereitstellung eines API (Application Programming Interface) für Bibliotheken im D-A-CH-Raum, ergänzt um ein Webinterface zur intellektuellen Prüfung nicht-verifizierbarer Daten

Datenformat des API	JSON	basierend auf der Dublin Core Metadata Initiative (DCMI)
Erforderliche Datenfelder	<p>title</p> <p>ispartof [Gesamttitel für Bände von Mehrteilern]</p> <p>creator</p> <p>date</p> <p>publisher</p> <p>identifizier [spezifisch für gebenden Institution]</p>	<p>„title“ zwingend erforderlich</p> <p>„ispartof“ ist notwendig für Einzelbände ohne eigenen Titel</p> <p>weitere Datenfelder sind möglich</p> <p>empfohlen</p>
Was passiert mit den Daten?	<p>automatisierte Identifikation der Transliterationsnorm</p> <p>automatische Retransliteraton mit Ranking der Konversionsmöglichkeiten</p>	<p>heuristisches Verfahren</p> <p>Python-Tool; heuristisches Verfahren auf Grundlage der identifizieren Transliterationsnorm</p> <p>Ranking mit Hilfe hebräischer Wörterbücher, dem Cache verifizierter Formen und häufiger Abweichungen/Fehler</p>
	Verifikation	<p>automatisierter Abgleich der konvertierten Daten mit originalschriftlichen Daten der NLI / der UB JCS</p>

Datenausgabe	Datenausgabe in drei Kategorien: verified unverified error	verifizierte, konvertierte Daten werden mit den Abgleichsdaten überschrieben
	Angabe detaillierter diagnostischer Informationen und Empfehlung display search	Empfehlungen für der Übernahme der Daten in anzeigbare Datenfehler (display) oder nur in indexierte Kategorien gegeben (search) → verbessert bereits die Auffindbarkeit von Hebraica Bei nicht-verifizierten Konversionen kann für jedes Titelsubfeld die individuelle Empfehlung variieren.
Möglichkeiten zur Datenprüfung	Webinterface zur intellektuellen Prüfung	zur Überprüfung nicht-verifizierter Daten, wird in Kürze bereitgestellt
Anwendung	geeignet für hebräische Titel	für jiddische Titel ist die Umsetzung der automatischen Retransliteration für die nächsten Monate geplant

Dokumentationen

Die Entwicklung steht unter einer CC0-Lizenz und ist über das [GitHub der Universitätsbibliothek Frankfurt am Main](#) mit einer ausführlichen Dokumentation zugänglich.

- Konversionsverfahren [Deromanize](#)
- [Dokumentation](#)
- Aufruf des Revrit-[API](#)

Beispiele

	Datenausgabe des Typs verified		Datenausgabe des Typs unverified
type	"verified"	type	"unverified"
title	"{ha-} Zemer ha-ivri : po'eṭikah, musiḳah, hištoriyah, tarbut / 'orekhet ha-ḳovets Tamar Ṿolf-Monzon"	title	"Maṭeh Dan ve-Kuzari ḥelek sheni : yokhiaḥ ve-yoreh bi-te'anot sikhliyyot bi-re'ayot ḥazaḳot u-ve-moftim gedolim amitat she-be'al peh ha-meḳubelet me-R[abotenu] z[ikhronam] l[ivrakhah] ḥokhme ha-Mishnah ve-Shas"
isPartOf	"Biḳoret u-parshanut"		
_seriesFields	"036E/00"		
creator	"Ṿolf-Monzon, Tamar"	creator	"Nieto, David"
_creatorFields	"028C"	_creatorFields	"028A"
date		date	
0	2012	0	1780
		1	1779
		2	80
publisher	"Universiṭat Bar-Ilan, Ramat-Gan"	publisher	"Shpayer, Mets"
_publisherFields	"033A"	_publisherFields	"033A"
identifier	"728971356"	identifier	"083888616"

	Datenausgabe des Typs verified		Datenausgabe des Typs unverified
converted	" הזמר העברי : פואטיקה, מוסיקה, היסטוריה, תרבות / עורכת הקובץ תמר וולף-מונזון"	converted	" מטה דן וכוזרי חלק שני : יוכיח ויורה בטענות סכליות בראיות חזקות ובמופתים גדולים אמיתת שבעל פה המקובלת מרז חכמי המשנה וש"ס"
matched_title	" הזמר העברי : פואטיקה, מוסיקה, היסטוריה, תרבות / עורכת הקובץ: תמר וולף-מונזון"	top_query_result	" מטה דן וכוזרי שני : ... יוכיח ויורה בטענות שכליות ... להכריח אמיתות תורה שבעל פה המקובלת מרז"ל חכמי המשנה ... והש"ס ע"ד הכוזרי הראשון ... / חברו הרב ... דוד ניטו"
link	" https://www.nli.org.il/en/books/NNL_ALEPH003454760/NLI "	link	" https://www.nli.org.il/en/books/NNL_ALEPH001243896/NLI "
diff	0		
diagnostic_info		diagnostic_info	
main_title		main_title	
standard	"New DIN 31631"	standard	"New DIN 31631"
foreign_tokens	false	foreign_tokens	false
transliteration_tokens	true	transliteration_tokens	true
fully_converted	true	fully_converted	true
all_cached	true	all_cached	true
all_recognized	true	all_recognized	true
subtitle		subtitle	
standard	"New DIN 31631"	standard	"New DIN 31631"
foreign_tokens	false	foreign_tokens	false
transliteration_tokens	true	transliteration_tokens	true
fully_converted	true	fully_converted	true
all_cached	true	all_cached	false
all_recognized	true	all_recognized	true
responsibility		responsibility	
standard	"New DIN 31631"	standard	"New DIN 31631"
foreign_tokens	false	foreign_tokens	false
transliteration_tokens	true	transliteration_tokens	true
fully_converted	true	fully_converted	false
all_cached	false	all_cached	false
all_recognized	false	all_recognized	false
recommendation		recommendation	
display	"matched_title"	display	[]
search	[]	search	
		0	"main_title"
		1	"subtitle"

Kontakt

Fachinformationsdienst Jüdische Studien
Goethe-Universität Frankfurt am Main
Bockenheimer Landstraße 134-138
60325 Frankfurt am Main
E-Mail: info@jewishstudies.de
Web: www.jewishstudies.de