

Webarchiv Zentralasien

Relevanz – Strukturen – Arbeitsschritte – Qualitätssicherung

Inhaltsverzeichnis

1. EINLEITUNG

- 1.1 WAS WURDE IN DER ERSTEN FÖRDERPHASE DES FID (2019-2021) GELEISTET?

2. VORGEHENSWEISE BEI AUSWAHL, SPIEGELUNG UND KATALOGISIERUNG

- 2.2 AUSWAHL DER WEBSEITEN FÜR DAS WEBARCHIV
- 2.3 AUSWAHLKRITERIEN FÜR URL VON WEBSEITEN, DIE AN DAS CRL GEMELDET WERDEN:

3. QUALITÄTSKONTROLLE AM BEISPIEL AUSGEWÄHLTER WEBSEITEN

- 3.1 BEISPIEL: TAGESZEITUNGEN
- 3.2 TAGESZEITUNG VON XINJIANG IN UIGURISCHER SPRACHE:
- 3.3 ÜBERSICHT GESPEICHERTE ZEITUNGSWEBSEITEN SINKIANG/KASACHSTAN:

4. KATALOGISIERUNG UND ERSCHLIEßUNG VON WEBSEITEN

- 4.1 BEISPIEL EINER REGIERUNGSWEBSEITE AUS SINKIANG:
- 4.2 DIE UNTERSEITEN HEIßEN:

5. LITERATUR IN AUSWAHL

6. ANHANG

1. Einleitung

„Die Webarchivierung hat sich seit Ende der neunziger Jahre zu einem neuen Handlungsfeld für Bibliotheken entwickelt.“ (Beinert et. al.: 2014, S. 291). Aus diesem Grunde sind an deutschen Forschungseinrichtungen und Bibliotheken Webarchive mit verschiedenen Schwerpunkten aufgebaut worden, an der SUB Göttingen im FID Zentralasien mit dem Schwerpunkt Zentralasien.

Die SUB Göttingen arbeitet eng mit dem Centre for Research Libraries in Chicago, USA zusammen. Eine ähnliche Politik verfolgt die Bayerische Staatsbibliothek (BSB), die ebenfalls eng mit wissenschaftlichen Einrichtungen in den USA zusammenarbeitet. Ein Beispiel ist das Archiv „Belarus Crisis“, das die BSB innerhalb des FID Ost-, Ostmittel- und Südosteuropa aufbaut.¹ Ziel ist immer die rechtskonforme, dauerhafte und offen zugängliche Speicherung der Internetquellen im Internet Archive, oft über die Dienste von Archive.It.

Webseiten sind meist nur für kurze Zeit im Netz verfügbar und verändern ihre Struktur und ihren Inhalt stetig. Im Raum Zentralasien sind die Webseiten zudem besonders durch Zensur bedroht. Deshalb hat der FID Zentralasien die Archivierung der Webseiten aus diesen Regionen zu seiner Aufgabe gemacht.

Zum Archivierungsziel im FID Zentralasien gehören die Webseiten in Turksprachen der betreuten Region (uigurisch, kasachisch, usbekisch, kirgisisch, turkmenisch, tatarisch, jakutisch etc.) und in mongolischen Sprachen. Da die Archivierung, Katalogisierung und Erschließung der Webseiten in Turksprachen aus der Uigurischen Autonomen Region Xinjiang der Volksrepublik China aufgrund der politischen Situation in diesem Gebiet in den letzten Jahren im Fokus der Arbeit standen und stehen, wird auf die Webseiten in den Turksprachen des zentralasiatischen Teiles Chinas unten besonders eingegangen.

Es gibt in Xinjiang Webseiten in uigurischer, kasachischer und kirgisischer Sprache in arabischer Schrift. Hinzu kommen Webseiten in oiratischer (westmongolischer) Sprache und Schrift. Die überwiegende Mehrheit der Webseiten ist in uigurischer Sprache. Es gab bis 2005 etwa 4200 Webseiten in uigurischer Sprache (Cf. Aziz Isa Elkun (2011). Elkun (Äziz Äyza Älkün) hat in seinem genannten Artikel mit Hilfe von vielen Beispielen belegt, dass die chinesische Regierung seit dem 05.07.2009 auf Webseiten in uigurischer Sprache Einfluss genommen hat – seit 2018 im verstärkten

¹ Belarus Crisis : Fall 2020

https://archive-it.org/organizations/985?fc=meta_Subject%3ApoliticsAndElections&flimit=fmeta_Subject%3A15

Maße. Die Frage, wie viele Webseiten in uigurischer Sprache gegenwärtig in Xinjiang noch abrufbar sind, lässt sich nicht genau beantworten. Die Situation ist sehr volatil. Das Beispiel der Webseite des Chinesischen Nationalen Radios in uigurischer und kasachischer Sprache, die seit Ende Oktober 2020 nicht mehr abrufbar ist ² zeigt, dass Webseiten in Sprachen der Minderheiten aktuell bedroht sind. Von daher kommt dem Projekt „Webarchiv Zentralasien“ innerhalb des FID Zentralasien eine besondere Bedeutung zu.

Besonderes Augenmerk wurde nicht nur auf konventionelle Webseiten gelegt, sondern gezielt auf Webseiten von Tages- oder Wochenzeitungen, die oft ein ganzes Archiv älterer Ausgaben bereithalten (s.u.), und auf Webseiten, oft Webseiten der Exiluguren, die im Sinne eines Repositoriums, bzw. einer Datenbank Monographien, Zeitschriften, einzelne Artikel, Audio- und Videobeiträge, also eine ganze Online-Bibliothek in uigurischer Sprache bereithalten.³ Auf Struktur und Inhalt der Webseiten wird unten genauer eingegangen.

Da eine reine Archivierung des Materials in der Wayback Machine des Internet Archive für die Wissenschaft nur schwer gezielt nutzbar ist, wurden archivierte Webseiten in der ZDB katalogisiert, verlinkt und durch Schlagwörter und Beschreibungen auf Deutsch und Englisch inhaltlich erschlossen und überregional über den GVK suchbar gemacht. Eine Anleitung zur Recherche wurde auf der Homepage des FID Zentralasien zur Verfügung gestellt. <https://fid-cassib.de/webarchive-central-asia/>

1.1 Was wurde in der ersten Förderphase des FID (2019-2021) geleistet?

Im Folgenden werden die Schritte zum **Aufbau des Webarchivs Zentralasien** beschrieben.

Juni 2019: Einstellung eines uigurischen Muttersprachlers als Mitarbeiter im FID Zentralasien, zuständig für das Webarchiv Zentralasien.

Zweites Halbjahr 2019: Abstimmungsprozesse und Sammlung von Webseiten aus Xinjiang und Zentralasien:

Beitrittsverhandlungen mit dem **Centre for Research Libraries** und indirekt über das CRL mit dem **Archive.It**, beide USA.

² (https://web.archive.org/web/2020*/http://uycnr.com/ , URL in „Central Asian Web Archiving: https://wayback.archive-it.org/13340*/http://www.uycnr.com/)

³ Ein Beispiel: <https://www.uygur.com/> mit mehreren Unterseiten Monographien, z.B. :

<https://www.uygur.com/kutuphana/> Die Seite wurde bereits gespiegelt

https://web.archive.org/web/*/https://www.uygur.com/, jedoch bestehen noch (16.3.2021)

Probleme bei der Speicherung der PDFs. An einer Lösung arbeitet das CRL. Die Hauptseite ist in der ZDB formalerschlossen und suchbar im GVK unter dem Titel: Uyğur tori ئۇيغۇر تورى

Seit **1. November 2019** ist die SUB Göttingen Mitglied beim CRL. Das CRL hat Ende 2019 mit dem Archive.It Verhandlungen zum Aufbau einer Collection „**Central Asian Web Archiving**“ geführt. Diese Sammlung steht ab Januar 2020 über die Wayback Machine des Internet Archive zur Verfügung. <https://archive-it.org/collections/13340>

Juni bis Dezember 2019: Sammlung von über 700 URL von uigurischen, kasachischen, kirgisischen und mongolischen Webseiten in den autochthonen Sprachen der Regionen Sinkiang, Innere Mongolei (siehe unten PDF-Anhang), teilweise auch von exilugurischen Seiten oder der Uiguren in Kasachstan und Kirgistan.

Der Mitarbeiter im FID nahm im Herbst 2019 an einer aufwendigen Schulung zur originalschriftlichen Katalogisierung und verbalen Erschließung ausländischer Webseiten in der ZDB teil.

Erstes Halbjahr 2020: Schulungsphase 2, erste Test Crawls durch Archive.It und Nachbearbeitung der Erst-Crawls durch die SUB Göttingen

URL ausgewählter uigurischer, kasachischer, mongolischer Webseiten werden dem CRL geschickt. Diese entscheidet über die Relevanz und bittet ab Februar 2020 das Archive.It diese zu spiegeln. Erste Test-Crawls liegen ab März 2020 vor und werden von der CRL und einem Mitarbeiter des FID Zentralasien analysiert. Entsprechend dem Abkommen zwischen dem CRL und dem Archive.It für das „Central Asian Web Archiving“ dürfen die Crawls für 2020 das Datenvolumen von 1.0 Terabyte nicht überschreiten. Hierfür zahlt die CRL an das Archive.It 12.000 US \$.

Zwischen CRL und der SUB Göttingen fanden im April und Juni 2020 zwei Video Konferenzen zur Analyse der Test Crawls und zum weiteren Vorgehen statt.

Es wurde festgestellt, daß Hyperlinks im Allgemeinen und insbesondere zu eingebetteten PDFs und Audio/Video Dateien zunächst bei den Test Crawls nicht aktiv gespeichert wurden. Dies konnte in der Folge durch das CRL in Zusammenarbeit mit dem Archive.It teilweise gelöst werden. So konnten etwa die PDFs vieler Tageszeitungsausgaben aus Sinkiang archiviert werden. Bei fast jeder neu gespeicherten Webseite treten jedoch wieder die alten Probleme auf, die individuell angesprochen und gelöst werden müssen. Die Webseiten müssen auch später regelmäßig daraufhin überprüft werden, ob die Probleme behoben werden konnten.

Ältere Crawls (vor 2020) uigurischer Seiten in der Wayback Machine weisen ebenfalls tote Hyperlinks auf. Durch die enge Zusammenarbeit von FID Zentralasien, CRL und Archive.It konnte die Situation bei Crawls solcher Seiten seit 2020 signifikant verbessert werden.

Nach einer ersten Testphase werden die gespiegelten Seiten dauerhaft im „Central Asian Web Archiving“ und anschließend in der Wayback Machine gespeichert und sind dort offen zugänglich.

Zweites Halbjahr 2020/Januar bis März 2021: Freigabe der gesammelten Webseiten zur Archivierung, sowie Katalogisierung und Erschließung der Webseiten

Um das Webarchiv qualitativ gut gestalten zu können, haben wir die gesammelten Webseiten ausführlich beobachtet und nach inhaltlicher, wissenschaftlicher und politischer Relevanz dieser Webseiten sortiert und für die Aufnahme in das Webarchiv eine engere Auswahl getroffen. Anschließend wurden gespeicherte Webseiten in der ZDB formal und inhaltlich erschlossen. Die Formalerschließung geschieht in Originalschrift mit international gebräuchlichen Umschriften.

2. Vorgehensweise bei Auswahl, Spiegelung und Katalogisierung

2.2 Auswahl der Webseiten für das Webarchiv

Bisher wurden insgesamt 109 (Stand 26.3.21) Webseiten in uigurischer, kasachischer, kirgisischer und mongolischer Sprache in das Webarchiv „Central Asian Web Archiving“ aufgenommen. Das jährliche Datenvolumen der gespeicherten zentralasiatischen Webseiten, bzw. der Crawls ist auf 1.0 Terabyte begrenzt. Der Auswahl liegen folgenden Kriterien zu Grunde:

Die zu spiegelnden Webseiten wurden unter linguistischen und gesellschaftswissenschaftlichen Gesichtspunkten ausgewählt.

Die Themenbereiche der gespiegelten Webseiten decken Bildung und Erziehung, uigurische und kasachische Geschichte und Kultur, Sprache und Literatur, sowie traditionelle Ethnomedizin und andere Bereiche wie Wirtschaft, Jura etc. ab, letztere vor allem aus sprachlichen Gesichtspunkten. Darunter gibt es viele Regierungswebseiten, Unternehmenswebseiten und persönliche Blogs⁴ usw., alle in den autochthonen Sprachen der Region.

Um die Vielfalt der uigurischen, kasachischen, kirgisischen, mongolischen Sprache wiederzuspiegeln, wurden bewußt einzelne Webseiten aufgenommen, die inhaltlich aus kultur- und gesellschaftswissenschaftlicher Sicht weniger relevant sind. Sie sind für den Sprachwissenschaftler wichtig. Exilugurische Webseiten wurden nur berücksichtigt, wenn kulturelle Aspekte der Uiguren enthalten sind:

⁴ Blog, bzw. Weblog in uigurischer Sprache (uig) wird im GVK so gesucht: aiw weblog spr uig. Für andere Sprachen: kaz (Kasachisch), mon (Mongolisch), kir (Kirgisisch).

2.3 Auswahlkriterien für URL von Webseiten, die an das CRL gemeldet werden:

- Regierungswebseiten in uigurischer, kasachischer, kirgisischer und mongolischer Sprache werden in das Webarchiv aufgenommen, wenn auf diesen Webseiten relevante Informationen über die Politik, Wirtschaft, Gesellschaft und Kultur dieser Region publiziert werden. Die Zahl dieser Webseiten nimmt ab. So gab noch vor zwei Jahren die Regierungswebseite des Kirgisischen Autonomen Bezirks Qizilsu in Atush in kirgisischer Sprache. Gegenwärtig ist sie nicht abrufbar.
- Unternehmenswebseiten in uigurischer Sprache aus Xinjiang (Gesundheit, Wirtschafts- und Finanzwesen, IT- und Smartphone-Technologie, Lebensmittel und Kochkunst, Kosmetik etc-) werden selektiv gespiegelt. Linguistische Forschungen zum Aussterben von Kulturen und Sprachen haben gezeigt, dass die Terminologien und Wörter in diesen Bereichen zuerst verschwinden. Darum ist die Archivierung solcher Webseiten auch für die Nachwelt und Wissenschaft, insbesondere die Sprachwissenschaft von hoher Relevanz.
- einige Informationsportale, Jobportale und Nachrichtenportale in uigurischer Sprache aus Xinjiang, die durch Nichtregierungsorganisationen und Personen erstellt und gepflegt werden, werden in das Webarchiv aufgenommen, weil auf diesen Webseiten wichtige Informationen über die soziale Situation, gesellschaftliche und politische Strukturen in Xinjiang subtil zum Ausdruck kommen und somit für die Nachwelt zum Verständnis der kontemporären Gesellschaft relevant sind.
- Webverzeichnisse in uigurischer Sprache aus Xinjiang werden selektiv ausgewählt und einige davon in das Webarchiv aufgenommen, weil sich hier Listen und Hinweise auf uigurische Webseiten finden, die bereits früher gelöscht wurden. Von daher können die Webverzeichnisse wichtige Auskünfte den kommenden Generationen über die Entwicklung der Internetkultur der Uiguren geben.
- Webseiten wissenschaftlicher Organisationen in und außerhalb Xinjiangs, die sich mit der uigurischen Kultur, Sprache und Literatur in uigurischer Sprache beschäftigen, werden in das Webarchiv aufgenommen. Dies gilt auch für andere autochthone Völker und Sprachen der Region.
- Nachrichtenwebseiten im Exil und Webseiten der politischen und kulturellen Organisationen der Exiliguren in uigurischer Sprache werden in das Webarchiv selektiv aufgenommen, sofern auf diesen Webseiten wichtige Informationen über die Kultur der (Exil-)Uiguren publiziert werden.
- Webseiten über den Islam und mit islamischen Inhalten in uigurischer Sprache der Exiliguren werden selektiv aufgenommen.
- persönliche Blogs werden selektiv ausgewählt und nur einige davon werden in das Webarchiv aufgenommen, wenn Beiträge in diesen Blogs für die Wissenschaft und zur Dokumentierung der uigurischen, kasachischen, kirgisischen etc. Kultur von der Bedeutung sind.

- einige wenige Filmportale werden in das Webarchiv aufgenommen, sofern sich auf diesen wertvolle Kulturzeugnisse der Uiguren, Kasachen etc. (Filme, Bilder, Kultursendungen, etc.) gespeichert finden.

3. Qualitätskontrolle am Beispiel ausgewählter Webseiten

URLs werden an das CRL geschickt und gegebenenfalls von dem CRL an Archive.It weiter gemeldet. Archive.It spiegelt dann die Seiten. Der Mitarbeiter des FID Zentralasien kontrolliert, ob die Seite vollständig gespeichert wurde und meldet mögliche Probleme mit einer Analyse des Erstcrawls an das CRL. Auch Folgebrawls müssen ständig auf ihre Qualität überprüft werden.

Dies soll hier am Beispiel der **Webseite der Parteikommission der Uigurischen Autonomen Region Xinjiang** kurz geschildert werden. Die URL dieser gespeicherten Webseite lautet: <https://wayback.archive-it.org/13340/20200508214909/http://uy.ts.cn/>

Anfangs wurden viele Unterkategorien und Beiträge, besonders Videobeiträge dieser Webseite nicht gespeichert, bzw. die Hyperlinks waren in der Speicherung nicht aktiv. Darum hat der FID-Mitarbeiter jede einzelne Kategorie dieser Webseite überprüft und die Probleme an das CRL gemeldet. Dadurch konnten in den Folgebrawls die meisten Probleme gelöst werden.

Besonders bei der Spiegelung der Webseiten der **Tageszeitungen aus Xinjiang** gab es zunächst Probleme: Die Tagesausgabe der jeweiligen Zeitung wird meist als PDF auf der Webseite kostenlos angeboten. Jedoch wurden die Hyperlinks zu den PDFs zunächst nicht oder nicht korrekt gespeichert. Dafür kann ein Beispiel genannt werden:

„Die Tageszeitung von Xinjiang“ (Chin. Xinjiang Ribao) erscheint in chinesischer, kasachischer, uigurischer und oiratischer (westmongolischer) Sprache. Der FID konzentriert sich hierbei auf die letzten drei Sprachen. Die Ausgaben in den verschiedenen Sprachen unterscheiden sich inhaltlich teilweise. Es finden sich PDF-Ausgaben auf der Webseite dieser Zeitung in den oben erwähnten Sprachen. URL der uigurischen Version:

<https://wayback.archive-it.org/13340/20200508214811/http://uyghur.xjdaily.com/> Link zur PDF-Version, die meist mit der Druckversion identisch ist:

<https://wayback.archive-it.org/13340/20200527210321/http://uy.xjdaily.com.cn/shuzibao/index.html>.

Zu Beginn der Archivierung gab es Probleme mit der Speicherung der PDF-Ausgaben dieser Zeitung in den oben erwähnten drei Sprachen. Das Problem wurde vom Mitarbeiter des FID an das CRL gemeldet. Das Problem wurde gelöst. Diese

Zeitungswebseiten sind besonders wertvoll, da sie Zugang auch zu vergangenen Ausgaben, oft über Jahre retrospektiv gewähren. Hier wurde also ein ganzes **Zeitungsbibliothek** mitgespeichert. Unten werden die Zeitungswebseiten ausführlicher beschrieben. Es wurden über 12.000 Zeitungssseiten (Stand 14.3.2021) als PDF oder Bild gespeichert.

3.1 Beispiel: Tageszeitungen

Es wurden bislang acht Tageszeitungen aus Sinkiang in den autochthonen Sprachen der Region im Central Asian Web Archiving gespeichert. Hinzu kommt die uigurische Wochenzeitung aus Kasachstan.

Insgesamt wurden ca. 12.000 Zeitungssseiten als PDF oder Bilddatei gespeichert (Stand 26.2.2021). Bei zwei Zeitungen gibt es noch Probleme bei der Speicherung der PDFs, die aber vermutlich gelöst werden können. Viele Zeitungswebseiten enthalten Archive älterer Ausgaben, die viele Jahre zurückreichen. Es kommt vor, dass Archive älterer Ausgaben auf der aktuellen Webseite gelöscht wurden, aber im Central Asian Web Archiving gespeichert bleiben.

Im Folgenden werden Beispiele für Zeitungswebseiten aus Xinjiang anführen.

3.2 Tageszeitung von Xinjiang in uigurischer Sprache⁵:

Die Webseite der Tageszeitung von Xinjiang (chin. Xinjiang Ribao) in uigurischer Sprache ist im Webarchiv gespeichert. Aus der chinesischen Version dieser Webseite kommt die Information, dass die uigurische, kasachische und mongolische Versionen dieser Zeitungswebseite sich gegenwärtig im Umbau befinden d.h. sie ist immer wieder nicht erreichbar. Gegenwärtig (Stand 26.03.2021) kann man nur die PDF-Ausgaben der Zeitungen in den jeweiligen Sprachen aufrufen (<http://uyghur.xjdaily.com/>). Der Link zu der uigurischen Webseite der Zeitung in unserem Webarchiv lautet:

https://wayback.archive-it.org/13340/*/http://uyghur.xjdaily.com/

Es gibt die Startseite. Hier sind sowohl vier Sprachen ganz oben auswählbar und darunter zudem acht Unterseiten. Die Startseite zeigt die aktuellen Druckausgaben. Die Unterseiten verweisen auf folgenden Kategorien:

⁵ Zeitungen in einer bestimmten Sprache werden im GVK so gesucht: Uigurische Zeitungen: aiw zeitung spr uig ; Kasachische Zeitungen: aiw zeitung spr kaz ; Mongolische Zeitungen: aiw zeitung spr mon ;

1. Startseite

<https://wayback.archive-it.org/13340/20200508214811/http://uyghur.xjdaily.com/>

Die acht Unterseiten

- 1.1 Muhim xäwärlär (dt. „wichtige Nachrichten“), URL:
<https://wayback.archiveit.org/13340/20200527205908/http://uy.xjdaily.com.cn/xw/yw/index.shtml>
- 1.2 Mämlikät Xäwärliri (dt. „Inlandsnachrichten“), URL:
<https://wayback.archiveit.org/13340/20200527210003/http://uy.xjdaily.com.cn/xw/gnxw/index.shtml>
- 1.3 Šinjang xäwärliri (dt. „Nachrichten zur Region Xinjiang“), URL:
<https://wayback.archiveit.org/13340/20200527210021/http://uy.xjdaily.com.cn/xw/xjhw/index.shtml>
- 1.4 Jaylarda (dt. „Lokales“),
<https://wayback.archiveit.org/13340/20200527210037/http://uy.xjdaily.com.cn/xw/dzxw/index.shtml>
- 1.5 Obzor (dt. „Kommentar“), URL:
<https://wayback.archiveit.org/13340/20200527210054/http://uy.xjdaily.com.cn/xx/pl/index.shtml>
- 1.6 Öginiš (dt. „Lernen“), URL:
<https://wayback.archiveit.org/13340/20200527210120/http://uy.xjdaily.com.cn/xx/index.shtml>
- 1.7 Sayahät (dt. „Tourismus“), URL:
<https://wayback.archiveit.org/13340/20200527210227/http://uy.xjdaily.com.cn/xw/lyxw/index.shtml>
- 1.8 Sipirliq gezit (dt. „digitale Zeitung“ / PDF-Zeitung), URL:
<https://wayback.archiveit.org/13340/20200527210321/http://uy.xjdaily.com.cn/shuzibao/index.html>

3.3 Übersicht gespeicherte Zeitungswebseiten Sinkiang/Kasachstan

Nr.	Webseite Zeitung Name/Link (Hauptseite)	Anzahl archivierter Seiten (PDF oder Bild)	Unterseiten ja/nein Beispiel Link	Anzahl archivierter Unter- seiten
1	„Šynjang gazeti“ Tageszeitung von Xinjiang in kasachischer Sprache / https://wayback.archive- it.org/13340/20210219185334 /http://kazakh.xjdaily.com.cn/	1432 (PDF-Seiten)		
2	„Šinjang geziti“ Tageszeitung von Xinjiang in uigurischer Sprache / https://wayback.archive- it.org/13340/20210212193834 /http://uy.xjdaily.com.cn/	1920 (PDF-Seiten)	https://wayback .archive- it.org/13340/20 200807184218 /http://uy.xjdaily .com.cn/xw/yw/ index.shtml	8
3	„Šinĵiyang-giyin ödör-in sonin“ Tageszeitung von Xinjiang in oiratisch- mongolischer Sprache / https://wayback.archive- it.org/13340/20210212193834 /http://mongol.xjdaily.com/ (in oiratischer und mongolischer Schrift)	536 (PDF-Seiten)	https://wayback .archive- it.org/13340/20 201016182747 /http://mongol.x jdaily.com/126/i ndex.html	14
4	„Altay gazeti“ ⁶ / https://wayback.archive- it.org/13340/20200908145717 /http://kazakh.altxw.com/news /altrb/index.shtml			7

⁶ „Die Tageszeitung von Altay“ ist eine kasachische Zeitung und erscheint jeden Tag mit acht Seiten außer an Wochenenden und Feiertagen. Die Ausgaben seit Februar 2020 sind auf dieser Webseite in Form von Bilddateien vorhanden. Wir suchen jetzt mit der CRL zusammen eine Lösung, alle diese Bilddateien zu speichern. Die Nachrichtenwebseite des Kasachischen Autonomen Bezirks hat sieben Unterseiten, die vor kurzem an die CRL gesendet wurden und bald gespeichert werden, vgl.: URL: <http://kazakh.altxw.com/>, zuletzt abgerufen am 26.02.2021.

5	„Ili geziti“ ⁷ (uigurisch) / „Ile gazeti“ (kasachisch) / https://wayback.archive-it.org/13340/20210212182506/http://epaper.ylxw.com.cn/user/y/lrb.html			18
6	„Tarbağatay gazeti“ ⁸ / https://wayback.archive-it.org/13340/20210212182504/http://www.kztcxw.cc/category/58_1.html	7604 (PDF-Seiten)	https://wayback.archive-it.org/13340/20200918184340/http://www.kztcxw.cc/category/18.html	18
7	„Uygur avazi“ ⁹ (Wochenzeitung der Uiguren in Kasachstan in uigurisch-kyrillischer Schrift) / https://wayback.archive-it.org/13340*/http://uyguravazi.kazgazeta.kz/	296 (PDF-Seiten)	https://wayback.archive-it.org/13340/20210212191248/http://uyguravazi.kazgazeta.kz/rubric/118/news	9

⁷ „Die Tageszeitung von Ili“ erscheint in uigurischer und kasachischer Sprache an Wochentagen mit Ausnahmen von Feiertage (4 bis 8 Seiten). Die PDF-Ausgaben dieser Zeitung seit Anfang Januar 2017 in diesen beiden Sprachen stehen auf der Nachrichtenwebseite des Bezirks von Ili kostenlos zur Verfügung. Wir haben mit der Speicherung dieser Dateien technische Schwierigkeiten. Die CRL bemüht sich jetzt, das Problem zu lösen. Diese Nachrichten- und Zeitungswebseite in kasachischer Sprache hat insgesamt 18 Unterseiten, die vor kurzem an die CRL gesendet worden sind und bald gespeichert werden können, vgl. URL: <http://kazakh.ylxw.com.cn/>), zuletzt abgerufen am 26.02.2021.

⁸ Die PDF-Ausgaben der „Tageszeitung von Tarbağatay“ in kasachischer Sprache sind seit 01.11.2013 mit Ausnahmen der Ausgaben zwischen Mitte September 2017 und Mitte Dezember 2017 lückenlos in unserem Webarchiv gespeichert. Die erste Ausgabe der PDF-Version dieser Zeitung ist am 01.11.2013 nachgewiesen: vgl. URL: https://wayback.archive-it.org/13340/20210212191652/http://www.kztcxw.cc/category/58_60.html. Die letzte Ausgabe ist am 10.02.2021 nachgewiesen: vgl. URL: https://wayback.archive-it.org/13340/20210212182504/http://www.kztcxw.cc/category/58_1.html, zuletzt abgerufen am 26.02.2021.

⁹ Diese Zeitung erscheint normalerweise wöchentlich in acht Seiten.

8	„Xälq tori“ ¹⁰ (Volkszeitung auf Uigurisch)/ https://wayback.archive-it.org/13340/20210212182642/http://uyghur.people.com.cn/	Nur Online Ausgabe	https://wayback.archive-it.org/13340/20210212195201/http://uyghur.people.com.cn/155988/309877/index.html	20
9	„Xalıq torabı“ ¹¹ (Volkszeitung auf Kasachisch)/ https://wayback.archive-it.org/13340/20210212182517/http://kazakh.people.com.cn/ (in kasachisch-arabischer Schrift)	Nur online Ausgabe	https://wayback.archive-it.org/13340/20210212210127/http://kazakh.people.com.cn/306608/index.html	21

4. Katalogisierung und Erschließung von Webseiten

Einzelne zentralasiatische Webseiten wurden bereits vor 2019 in der Wayback Machine des Internet Archive gespeichert. Crawls für diese Seiten lagen oft viele Monate auseinander. Hyperlinks wurden oft nicht gespeichert. Die gespeicherten Seiten waren kaum recherchierbar, sofern die URL nicht bekannt war.

Um eine gezielte thematische Suche nach zentralasiatischen Webseiten im Internet Archive zu ermöglichen, wurden bislang rund 500 zentralasiatische und exilugurische Webseiten mit Unterseiten in der ZDB katalogisiert, erschlossen und in englischer und

¹⁰ Online-Version der chinesischen Volkszeitung (chin. Renmin Ribao). Die „Volkszeitung“ gilt in China als eine Parteizeitung der Kommunistischen Partei Chinas. Die "Volkszeitung" hat keine gedruckte Ausgabe in uigurischer Sprache. Diese Zeitungsw Webseite hat verschiedene sprachliche Versionen. Darunter sind die Sprachen von sieben nationalen Minderheiten Chinas. Außerdem finden sich Versionen in neun ausländischen Sprachen. Diese Online-Zeitung ist in den oben erwähnten verschiedenen Sprachen mit unterschiedlichen Inhalten vorhanden.

¹¹ Online-Version der chinesischen Volkszeitung (chin. Renmin Ribao). Die "Volkszeitung" hat keine gedruckte Ausgabe in kasachischer Sprache. Diese Online-Zeitung in kasachischer Sprache ist sowohl in kasachisch-arabischer Schrift als auch in kyrillischer Schrift vorhanden: vgl. URL: <https://wayback.archive-it.org/13340/20210213123233/http://kazakh.people.com.cn/reold/kry/>, zuletzt abgerufen am 26.02.2021.

deutscher Sprache kurz beschrieben. Diese Katalogisate sind weltweit über den Gemeinsamen Verbundkatalog (GVK) zugänglich. Eine Anleitung zur Recherche hier: <https://fid-cassib.de/webarchive-central-asia/>

Bedingt durch die komplexe Struktur vieler Webseiten, wurden thematisch abweichende Unterseiten gesondert katalogisiert.

Jedes Katalogisat enthält u.a. die **folgenden Kategorien**, die auch einzeln recherchierbar sind:

- Erscheinungsdatum (nachgewiesen seit)
- Sprachcode
- Ländercode
- Erscheinungsort
- Titel der Webseite (ggf. Paralleltitel, abweichender Titel und frühere Haupttitel. Originalschriftliche Erfassung und Transliteration des Titels)
- Autoren/Körperschaften/Parteien (Originalschriftliche Erfassung und Transliteration)
- Spiegelungsdatum
- Schlagwörter (Entsprechend den Regeln der ZDB lassen sich für eine Titelaufnahme nur sechs Schlagwörter vergeben.)
- URL (aktuell zum Zeitpunkt der Katalogisierung, ggf. später aktualisiert)
- Link zur archivierten Webseite im Internet Archive/Central Asian Web Archiving
- Ggf. Zugriffsbeschränkung
- Inhaltliche Beschreibung (Es sind nur sechs Schlagwörter möglich. Daher wird im Feld 4207 die Webseite nochmals auf Englisch und Deutsch in gebundener Sprache beschrieben.)
- Schriftart

4.1 Beispiel einer Regierungswebseite aus Sinkiang:

Die Aufnahme der Eingangsw Webseite der Parteikommission Sinkiangs sieht im Katalogisierungssystem zunächst wie folgt aus:

Eingabe: 0007:10-02-20 Änderung: 0007:22-02-21 18:18:42 Status: 0007:10-02-20 **0500** Obxz

0501 Text**\$b**txt

0502 Computermedien**\$b**c

0503 Online-Ressource**\$b**cr

1100 2001\$[2001]-

1101 cr

1131 [!959344357!](#)Website [Ts1]

1140 ws

1500 /1uig

1505 \$erda

1700 /1XB-CN

2110 3007987-1

2240 ZDB:3007987-1

4000 \$T01\$ULatn%%Tiyānšan tori / Šinjang uyğur aptonom rayonluq xälq hökümiti axbarat işxanisi; Näšir hoquqi Šinjang yengi taratqular märkizigä täwä 4000

\$T01\$UArab%%شەنجاڭ ئۇيغۇر ئاپتونوم رايونلۇق خەلق ھۆكۈمىتى ئاخبارات تىيانشان تورى / شەنجاڭ ئۇيغۇر ئاپتونوم رايونلۇق خەلق ھۆكۈمىتى ئاخبارات تىيانشان

ئىشخانىسى; نەشىر ھوقۇقى شەنجاڭ يېڭى تاراتقۇلار مەركىزىگە تەۋە

4030 Ürümçi : Šinjang uyğur aptonom rayonluq xälq hökümiti axbarat işxanisi

4060 Online-Ressource

4085 =u <https://web.archive.org/web/20171221223050/http://uy.ts.cn:80=x H=z LF>

4085 =u https://wayback.archive-it.org/13340*/http://uy.ts.cn/=x H=z LF

4085 =u <http://uy.ts.cn/=x H=z LF>

4201 Gesehen am 24.02.2020

4207 Regierungswebseite der Parteikommission von Xinjiang und Nachrichtenwebseite der

Kommunistischen Partei Chinas in Xinjiang. Es werden Beiträge und Nachrichten zur Politik, Gesellschaft, Wirtschaft, Landwirtschaft, Kultur, Bildung und Landeskunde von Xinjiang publiziert. Diese Webseite ist in verschiedenen Sprachen Uigurisch, Kasachisch, Mongolisch und Chinesisch sowie Russisch und Englisch abrufbar. - Government website of the Party Commission of Xinjiang and news website of the Chinese Communist Party in Xinjiang. It publishes articles and news on politics, society, economy, agriculture, culture, education and regional studies of Xinjiang. This website is available in various languages Uyghur, Kazakh, Mongolian and Chinese as well as Russian and English.

4213 \$T01\$ULatn%%Haupttitel 2001-2018: Tängritag tori

4213 \$T01\$UArab%%Haupttitel 2001-2018: تەڭرىتاغ تورى

4221 In arabisch-uirgischer Schrift

5056 [FID]ZENTRALASIEN\$qDE-7

5100 [!040099377!](#)China [Tgz]

5101 [!040774600!](#)Sinkiang [Tg1]
5102 [!040319008!](#)Kommunistische Partei [Ts1]
5103 [!040614913!](#)Uiguren [Ts1]
5104 [!040465144!](#)Politik [Tsz]
5105 [!04136757X!](#)Partei politik [Ts1]
5109 (DE-7){DE-601}

7001 16-03-20 : x

4800 [!009004742!](#)341000-6 <7> Göttingen, Niedersächsische Staats- und
 Universitätsbibliothek Göttingen
7135 =u [http://opac.sub.uni-](http://opac.sub.uni-goettingen.de/DB=1/LNG=DU/CLK?IKT=12&TRM=1689724374)
[goettingen.de/DB=1/LNG=DU/CLK?IKT=12&TRM=1689724374](http://opac.sub.uni-goettingen.de/DB=1/LNG=DU/CLK?IKT=12&TRM=1689724374)
7800 666957975
7900 25-03-20 17:01:01.000
8001 %8g{cruuuuuuuuuuuuu}
8032 #1#
8034 Bestand s. OPAC (Link anklicken)
8597 h3609052775

Ein generelles Problem bei der Erschließung und Katalogisierung von Webseiten ist die Tatsache, dass sich die meisten Webseiten aus mehreren **Unterkategorien** zusammensetzen und entsprechend den Katalogisierungsregeln der ZDB jede einzelne Kategorie als ein eigenständiger Titel aufgefasst wird und eine eigene Titelaufnahme verlangt. Wenn man daher bei diesem Beispiel im Gemeinsamen Verbundkatalog (GVK) nach dieser Webseite unter *Tiyanšan tori* sucht, stößt man insgesamt auf 40 Titelaufnahmen: Eine Aufnahme für die Hauptseite und 39 weitere für Unterseiten:

4.2 Die Unterseiten heißen:

1. Muhim xäwär (dt. „wichtige Nachrichten“), 2. Šinjang (dt. „Xinjiang“), 3. Mämlikät (dt. „chinesisches Inland“), 4. Nüzäriyä (dt. „Theorie“), 5. Ekran¹²: Män wä mening wätinim (dt. Bildschirm: „Ich und mein Vaterland“), 5.1. Ekran: Män wä mening wätinim: Öginiš

¹² **Ekran**, ein russisches Fremdwort im Uigurischen, bedeutet **Bildschirm** und meint meist Videos. Wenn man oben auf den Reiter „Ekran“ klickt, öffnet sich eine Seite mit zahlreichen Videokategorien, die hier beispielhaft alle aufgeführt sind. „Bildschirm“ (Ekran) kann sinngemäß oft besser als **Mediathek** übersetzt werden.

dawamlaşmağa (dt. Bildschirm: „Ich und mein Vaterland: Lernen schreitet voran“), 5.2. Ekran: Män wä mening wätinim: Yengi дәwr, yengi nätijä, yengi sähipä (dt. Bildschirm: „Ich und mein Vaterland: Neue Ära, neue Leistung, neue Seite“), 5.3. Ekran: Män wä mening wätinim: Sältänätlik 70 yil küräščän yengi дәwr (dt. Bildschirm: „Ich und mein Vaterland: Die großartigen 70 Jahre – ein kämpferisches neues Zeitalter“), 5.4. Ekran: Män wä mening wätinim: Namratliqtin qutulduruş ötkiligä hujum qiliş jengidä qätiy ğälibä qilayli (dt. Bildschirm: „Ich und mein Vaterland: Lasst uns im Sturm-Kampf gegen die Hürde der Armut kämpfen und bestimmt siegen“), 5.5. Ekran: Män wä mening wätinim: Mädäniyättä Şinjangğa näp yätküzüş (dt. Bildschirm: „Ich und mein Vaterland: Im Bereich der Kultur der Region Xinjiang Gewinn bringen“), 5.6. Ekran: Män wä mening wätinim: Millätlär ittipaqlišip bir a'ilä kişiliridäk boluş (dt. Bildschirm: „Ich und mein Vaterland: Nationalitäten werden zu Mitgliedern wie Mitglieder einer Familie, indem sie sich vereinigen“), 5.7. Ekran: Män wä mening wätinim: Yaşliq küräş qiliş üçündür (Bildschirm: „Ich und mein Vaterland: Die Jugendzeit des Menschen ist für den Kampf da“), 5.8. Ekran: Män wä mening wätinim: Jungxua änäniwi mädäniyitini äwj alduruş (Bildschirm: „Ich und mein Vaterland: Die traditionelle Zhonghua¹³-Zivilisation in Xinjiang verbreiten), 6. Räsimlik yäşmä (dt. „Erzählen mit Bildern“), 7. Äsliy mäqsätni untumay, burçni ästä čing saqlaş (dt. „Die ursprünglichen Zielvorstellungen im Kopf behalten und die Mission beherzigen“), 8. Kündä bir altun jümlä (dt. An jedem Tag einen Satz lernen, der einen Goldwert hat“), 9. Wätänpärwärlük hesyati, küräş qilğučılar (dt. „Patriotismus und die Kämpfenden“), 10. Tiyanşan tori mäxsus sähipisi: Iradidin sepil hasil qilip, yuqumning aldini eliş - tizginläş, tosüp zärbä beriş jengidä ğälibä qilayli (dt. Spezielles Forum der Webseite Tiyanşan: „Lasst uns aus unserem Willen eine feste Verteidigungsmauer bauen und lasst uns dadurch im Bereich des Kampfes in der Vorbeugung gegen eine Ansteckung, in den Kontrollmaßnahmen gegen die Infektion und im Angriff gegen die Infektion den Sieg erringen“), 11. Sältänätlik 70 yil küräščän yengi дәwr (dt. „Die großartigen 70 Jahre: Kämpferische neue Ära“), 12. Şinjangni qalaymiqan qiliş xam xiyalida bolğan härqandaq suyiqäst härgiz mäqsitigä yetälmäydu; Şinjangdiki här millät xälq qätiy qoşulmaydu (dt. „Jede Verschwörung mit leeren Phantasien, die das Gebiet Xinjiang ins Chaos stürzen will, darf keinesfalls ihr Ziel erreichen - Alle Nationalitäten in Xinjiang sind entschlossen, dies nicht zuzulassen.“), 13. Şinjang yaxşı jay (dt. „Xinjiang ist ein guter Ort“), 14. Millätning güllinişi üçün: Qährimanlar, qurbanlar şäjärisi (dt. „Für die Belebung der Nation: Die Biographien und Lebensgeschichten von Helden und Opferbringenden“), 15. Şinjangdiki bayqaş (dt. „Entdeckungen in Xinjiang“), 16. Ätimiz teximu güzäl: Käspi maharät tälim - tärbiyä: Ümidkä tolğan güzäl turmuş (dt. „Unsere Zukunft wird noch schöner: Erziehung und Ausbildung in den beruflichen Kompetenzen: das schöne Leben voller Hoffnung“), 17. Güzäl yurt - makanimizni yaxşı

¹³ Zhonghua: chin. „Kultur der Mitte“, i.e. die han-chinesische Kultur. Der Begriff steht oft für China an sich.

qoğdayli (dt. „Lasst uns unsere schöne Heimat gut verteidigen“), 18. Namratliqtin qutulduroş ötkiligä hujum qiliş (dt. Der Kampf gegen die Schwierigkeiten in der Armutsbekämpfung“), 19. Ärayini biliş, älgä näp yätküzüš, älni mayil qiliş (dt. „Sich über die Gefühle und Meinungen der Bevölkerung informieren, der Bevölkerung Gewinn bringen, die Herzen der Bevölkerung für die Regierung gewinnen“), 20. Millätlär ittitaqlışip bir a`ilä kişiliridäk boluş (dt. „Nationalitäten werden zu Mitgliedern wie Familienmitglieder einer Familie, indem sie sich einigen“), 21. Bingtuän (dt. „Xinjiang Produktions- und Aufbau-Korps“), 22. Şinjangğa nişanlıq yardäm beriş (dt. „Dem Gebiet Xinjiang gezielt helfen“), 23. Ürümçi (dt. „Urumchi“), 24. Mädäniyät – sayahät (dt. „Kultur und Tourismus“), 25. Jaylarda (dt. „Berichte und Nachrichten aus den Bezirken und Städten in Xinjiang“), 26. Pän - ma`arip (dt. „Bildungs- und Erziehungswesen in Xinjiang“), 27. Yeza igiliki (dt. „Landwirtschaft in Xinjiang“), 28. Täntärbiyä (dt. „Sport“), 29. Turmuş luğiti (dt. „Lebenswelt“), 30. Şinjang näzäriyä tori (dt. Spezielles Forum von Xinjiang für politische Theorie).

Die uigurischen Titel der oben angegebenen Unterseiten sind einzeln im GVK recherchierbar.

Bei der Katalogisierung kommt oft hinzu: Auch wenn der Ländercode, Erscheinungsdatum und Körperschaft(en) dieser Webseite aus den Informationen, die auf dieser Webseite zu finden sind, ermittelt werden können, müssen später noch das Feld 4207 und die Felder zwischen 5100 und 5105 ausgefüllt werden, nachdem man sich einen Überblick über den Inhalt aller Kategorien dieser Webseite gemacht hat. Es handelt sich bei dem Feld 4207 im Rahmen einer ZDB-Aufnahme um Angaben zum Inhalt elektronischer Ressourcen, während sich die Felder zwischen 5100 und 5105 aus Schlagwörtern zusammensetzen und sie damit auch den Inhalt des betreffenden Mediums mittels der Beschlagwortung erschließen.

In vielen Fällen gibt es auf den jeweiligen Webseiten keinerlei Informationen darüber, ob der Haupttitel sich im Laufe der Jahre geändert hat und wann dies möglicherweise geschah. Aber auch **Titelwechsel und alte Titel** müssen recherchiert und eingetragen werden. Diese Fragen sind für den Aufbau eines qualitativ hochwertigen Webarchivs und für eine gute Formal- und Sacherschließung von Webseiten von enormer Bedeutung. Hiervor tritt die reine Zahl katalogisierter Webseiten zurück.

Um das erwähnte Problem der **früheren Haupttitel** der Webseite der Parteikommission zu lösen, waren aufwendige Recherchen im Internet Archiv <https://archive.org/web/> notwendig. Unser Mitarbeiter im FID hat sich dort die Ausgaben der betreffenden Webseite aus früheren Jahren angeschaut und versucht, Antworten auf Fragen zu finden. Dies bedeutet einen erhöhten Zeit- und Arbeitsaufwand, um alte und neue Versionen einer Webseite zusammenzuführen.

Auf dieser Parteiwebseite gibt es auf vielen Unterseiten wie oben beschrieben **Videos**. Die meisten wurden bei der Spiegelung durch Archive.It mitgespeichert. Beim Aufruf **dauert es oft sehr lange**, bis die Videos geladen sind. Es folgen zwei Beispiele: Das erste Video heißt: „Baš šujining bir häptisi“ (dt. „Eine Woche im Leben des Generalsekretärs“):

<https://wayback.archiveit.org/13340/20210213120005/http://uy.ts.cn/system/2020/07/10/036355845.shtml>

Das zweite Beispiel ist das folgende Video: „Xotän wilayiti keriyä nahiyisidiki täsirlik hekayä“ (dt. „Eine beeindruckende Geschichte im Kreis Keriyä des Bezirks Hotan“):

<https://wayback.archiveit.org/13340/20210213140110/http://uy.ts.cn/system/2019/11/21/035972463.shtml>

Im Falle vieler uigurischer **Webseiten der Exiluiguren** ist es schwierig, den Ländercode, das Erscheinungsdatum, den Erscheinungsort zu bestimmen, da diese Angaben auf solchen Webseiten häufig aus unterschiedlichen Gründen fehlen.

Fragen zu Inhalten oder technischen Aspekten bleiben oft unbeantwortet.

5. Literatur in Auswahl

Äysa, Äziz Älkun (Azia Isa Elkun) (2011): „Uyğur tor bätçilikigä omumiy näzär: Uyğurlarda internetniñ qollinilişi – ärkin pikir bayan qilişniñ wastisi“ [Ein Gesamtüberblick über die uigurischen Webseiten: Die Verwendung des Internets durch die Uiguren als Mittel der freien Meinungsäußerung] (50 Seiten), dieser Aufsatz steht auf Uigurisch Online unter dem folgenden Link zur Verfügung: http://www.uyghurpen.org/uy/Uyghur_torbetchilikige_omumiy_nezer.pdf, zuletzt abgerufen am 26.03.2021.

Beinert, Tobias/Hagenah, Ulrich/Kugler, Anna (2014): „Es war einmal eine Website... - Kooperative Webarchivierung in der Praxis“, in: Das offene Bibliotheksjournal, H.1, S. 291-304, dieser Artikel ist unter dem folgenden Link abrufbar: <https://www.o-bib.de/article/view/2014H1S291-304>, zuletzt abgerufen am 26.03.2021.

Schumacher, Edgar (2016): „Webarchivierung an der SUB Hamburg: Ziele, Erfahrungen, Perspektiven“ (20 Seiten), dieser Artikel ist unter dem folgenden Link abzurufen: https://www.babs-muenchen.de/content/veranstaltung/2016-03-10/2016_03_10_schumacher.pdf, zuletzt gesehen am 26.03.2021.

Rauber, Andreas/Liegmann, Hans (2010): „Webarchivierung zur Langzeiterhaltung von Internet-Dokumenten“, in: nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Kap.17:88-Kap.17:103, Glückstadt: Verlag Werner Hülsbusch.

6. PDF-Anhang

Sammlung von URL uigurischer, kasachischer, kirgisischer, mongolischer Webseiten größtenteils aus China vom Mitarbeiter im FID 2019 erstellt.


Uyghur
language.pdf


Kazakh
language.pdf


Kyrgyz
language.pdf


Mongolian
language.pdf