

Building a Web Archive Central Asia¹ as Part of the Special Information Service for Central Asia and Siberia

Tahir Mutällip Qahiri

Introduction²

“Web archiving has developed into a new field of action for libraries since the late 1990s.” (Beinert et. al.: 2014, 291). For this reason, web archives with different focuses have been established at German research institutions and libraries. At the Göttingen State and University Library (SUB Göttingen) a so-called Web Archive Central Asia has been established.

The SUB Göttingen works closely together with the Centre for Research Libraries in Chicago, USA. A similar policy is pursued by the Bavarian State Library (BSB),

¹ I built the Web Archive Central Asia 2019 to 2021 and wrote a first version of this article for the CASSIB webpage which was then partly rewritten and presented by J. Reckel at the conference.

² All links quoted in this paper have been last accessed March 25th, 2022.

which also works together closely with academic institutions in the USA in archiving web sites. One example is the archive “Belarus Crisis”, which the BSB is setting up within the FID East, East-Central and Southeast Europe.³ The aim is always to store Internet sources in the Internet Archive in a legally compliant, permanent and openly accessible way, using the services of Archive.It.

Web pages are usually only available on the net for a short time and their structure and content are changing constantly. In the Central Asian region, websites are also particularly under threat by censorship and by political public unrest and regime change. For this reason, the FID Central Asia has made the archiving of websites from these regions its task.

The archival goal of the FID Central Asia includes the websites in Turkic languages of the region in question (Uyghur, Kazakh, Uzbek, Kirghiz, Turkmen, Tatar, Yakut etc.) and in Mongolian languages. Since the archiving, cataloguing and indexing of websites in Turkic languages from the Uyghur Autonomous Region of Xinjiang in the People’s Republic of China have been the focus of the work in recent years due to the political situation in this area, the websites in the Turkic languages of the Central Asian part of China will be dealt with in detail below.

There are websites in Uyghur, Kazakh and Kyrgyz languages in Arabic script in Xinjiang. In addition, there are websites in Oirat (Western Mongolian) language and script. The vast majority of websites are in Uyghur. There were about 4200 websites in Uyghur by 2005 (Cf. Aziz Isa Elkun (2011). Elkun (Äziz Äysa Älkün) has shown in his aforementioned article with the help of many examples that the Chinese government has been influencing Uyghur language websites since 05.07.2009 – since 2018 in an increased number of cases.

The question of how many websites in the Uyghur language are currently still accessible in Xinjiang cannot be answered precisely. The situation is very volatile. The example of the website of the Chinese National Radio in Uyghur and Kazakh, which has not been accessible since the end of October 2020,⁴ shows that websites in minority languages are currently under threat. For this reason, the project “Web Archive Central Asia” is of particular importance within the FID Central Asia.

Particular attention was paid not only to conventional websites, but specifically to websites of daily or weekly newspapers, which often hold an entire archive of older issues (see below), and to websites, often websites of exiled Uyghurs, which hold monographs, journals, individual articles, audio and video contributions, i.e.

³ Belarus Crisis: Fall 2020–2021 <https://archive-it.org/collections/14892>.

⁴ 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/.

an entire online library in the Uyghur language, in the sense of a repository or database.⁵ The structure and content of the websites will be discussed in more detail below.

Since it is difficult for scholars to make specific use of archived material in the Internet Archive's Wayback Machine, archived web pages in the Central Asian Web Archive were catalogued in the German Union Catalogue of Serials (ZDB), linked and their content indexed by keywords and descriptions in German and English, and made searchable beyond the region via the Common Network Catalogue (GVK). There are also keywords in English on the site of the Central Asian Web Archive of Archive.It itself. Instructions for users have been made available on the homepage of the FID Central Asia CASSIB. CASSIB stands for Central Asia – Siberia. <https://fid-cassib.de/webarchive-central-asia/>

Building a Web Archive for Central Asia (2019–2021)⁶

The following section describes the individual steps taken to build the Central Asian Web Archive.

June 2019: Hiring of a native Uyghur speaker as a staff member in the FID Central Asia project, responsible for the Central Asian Web Archive.

Second half of 2019: Coordination processes and collection of websites from Xinjiang and Central Asia: Negotiations with the Centre for Research Libraries and indirectly via CRL with Archive.It, both USA.

The SUB Göttingen has been a member of the CRL since November 1st, 2019. At the end of 2019, the CRL conducted negotiations with Archive.It to build a collection “**Central Asian Web Archiving**”, later renamed as “**Central Asian Web Archive**”. The content of this collection has been available via the Internet Archive's Wayback Machine from January 2020. <https://archive-it.org/collections/13340>

⁵ One example: <https://www.uygur.com/> with several sub-pages for monographs e.g. <https://www.uygur.com/kutuphana/> The page has been archived: https://web.archive.org/web/*/https://www.uygur.com/* Some PDF documents have not been saved correctly especially in early crawls. The main webpage ئۇيغۇر تۈرى has been catalogued in the ZDB and made more accessible by keywords. You can search it in the GVK (Common Library Network Catalogue) under ئۇيغۇر تۈرى

⁶ The more practical aspects of the Web Archive Central Asia, cataloguing and ways to search the archive have been addressed on the CASSIB webpage, which contains several important links. It also has a link to the original German version of this paper. This English version has been updated in March 2022. <https://fid-cassib.de/index.php?id=72&L=%201&cid=72>

June to December 2019: Collection of over 700 URL of Uyghur, Kazakh, Kyrgyz and Mongolian websites in the autochthonous languages of the regions of Sinkiang, Inner Mongolia, partly also of exiled Uyghur sites or of the Uyghurs in Kazakhstan and Kyrgyzstan

In autumn 2019, the FID staff member took part in an extensive training course on cataloguing and verbal indexing of foreign websites in the ZDB.⁷

First half of 2020: Training phase 2, first test crawls by Archive.It and post-processing of the initial crawls by the SUB Göttingen. The URL of selected Uyghur, Kazakh, Mongolian websites are sent to the CRL. The CRL decides on the relevance and asks Archive.It to crawl them from February 2020. The first test crawls have been available from March 2020 and have been analysed by the CRL and a staff member of the FID Central Asia. According to the agreement between the CRL and Archive.It for the “Central Asian Web Archiving”, the crawls for 2020 must not exceed the data volume of 1.0 terabyte. Later the data volume was increased. It was 2.8 TB for 2021. These had to be paid for from project money.

Video conferences were held between CRL and the SUB Göttingen in April and June 2020 to analyse the test crawls and to decide on the next steps.

It was found that hyperlinks in general, and in particular embedded PDFs and audio/video files, were not initially actively stored in the test crawls. This could subsequently be partially solved by the CRL in cooperation with Archive.It. For example, the PDFs of many daily newspaper editions from Sinkiang could be archived. However, with almost every newly saved website, the old problems reappear, which have to be addressed and solved individually. The web pages must also be regularly checked later to see whether the problems could be solved permanently. Older crawls (pre-2020) of Uyghur pages in the Wayback Machine usually have dead hyperlinks. Close cooperation between the FID Central Asia, CRL and Archive.It has significantly improved the situation for crawls of such pages since 2020.

After an initial test phase, the crawled pages will be permanently stored in the “Central Asian Web Archive” and in the Wayback Machine, where they will be openly accessible.

In order to enhance the quality of the web archive, the websites of the collected URL lists are constantly kept under observation and arranged according to their content and relevance for research. A shortlist of URLs for inclusion in the web archive was the first result. Subsequently, stored web pages were catalogued in the ZDB in terms of form and content. Formal cataloguing is done using the original script together with internationally accepted transcriptions.

⁷ I would like to thank Sonja Meier (SUB Göttingen) very much for her training offers on cataloguing digital media in the Zeitschriftendatenbank (ZDB) and her critical support of my work on cataloguing.

Procedure for Selection, Crawling and Cataloguing

Selection of Web Pages for the Web Archive

So far, a total of 112 (as of 10.1.22) major web pages in Uyghur, Kazakh, Kyrgyz and Mongolian have been included in the web archive “Central Asian Web Archive”. The selection is based on the following criteria: The websites to be crawled were selected under linguistic, cultural and sociological aspects.

The subject areas of the mirrored websites cover education and upbringing, Uyghur and Kazakh history and culture, language and literature, as well as traditional ethno-medicine and other areas such as economy, law, etc., the latter mainly from a linguistic point of view. Among them, there are many government websites, corporate websites and personal blogs⁸ etc., all in the autochthonous languages of the region.

In order to reflect the diversity of the Uyghur, Kazakh, Kyrgyz, Mongolian languages, individual websites were deliberately included that are less relevant in terms of content from a cultural and social science perspective. They are important for linguists. Exiled Uyghur websites have only been included if they contain cultural aspects of the Uyghurs.

Selection Criteria for URL of Websites Reported to the CRL

- Government websites in Uyghur, Kazakh, Kyrgyz and Mongolian language are included in the web archive if they contain relevant information about politics, economy, society and culture of this region. The number of these websites is decreasing. Three years ago, for example, the government website of the Kyrgyz Autonomous District of Qizilsu in Atush in the Kyrgyz language was still online. Currently (2020-2021) it is offline.
- Business websites in Uyghur from Xinjiang covering health, business and finance, IT and smart phone technology, food and cooking, cosmetics etc. are selectively crawled. Linguistic research into the extinction of cultures and languages has shown that terminologies and words in these fields disappear first. Therefore, the archiving of such websites is of great relevance for posterity and science, especially linguistics.
- Some information portals, job portals and news portals in Uyghur language from Xinjiang, which are created and maintained by non-governmental organisations and individuals, are included in the web archive because important information about the social situation, social and political structures in Xinjiang are subtly expressed and are therefore relevant for the understanding of contemporary society.

⁸ Blog, or weblog in Uyghur language (uig) is searched for in the GVK-catalogue like this: aiw weblog spr uig. For other languages exchange uig with: kaz (Kazakh), mon (Mongolian), kir (Kyrgyz) etc.

- Web directories in the Uyghur language from Xinjiang will be selectively included in the web archive because they contain lists and references to Uyghur websites that have been deleted earlier. From this point of view, the web directories can provide important information for future generations about the development of Uyghur internet culture.
- Websites of academic organisations in and outside of Xinjiang, which focus on Uyghur culture, language and literature, are included in the web archive. This also applies to other autochthonous peoples and languages of the region.
- News websites in exile and websites of the political and cultural organisations of the Uyghurs in exile in Uyghur language will be included in the web archive on a selective basis, if these websites contain important information about the culture of the (exiled) Uyghurs.
- Websites about Islam and with Islamic content in the Uyghur language of the exiled Uyghurs will be included selectively.
- Personal blogs will be chosen selectively and only some of them will be included in the web archive, if contributions in these blogs are useful for scholarly research and for documenting the Uyghur, Kazakh, Kyrgyz, Oirat culture.
- A few film portals will be included in the web archive, if technically and legally possible.

Quality Control Using the Example of Selected Websites

URLs are sent from Göttingen to the CRL and, if approved, reported by the CRL to Archive.It. Archive.It then initiates the crawling of these pages. The FID Central Asia staff member checks whether the page has been saved completely and reports possible problems to the CRL with an analysis of the initial crawl. Follow-up crawls must also be constantly checked for quality. Göttingen also decides if websites should be crawled weekly, monthly etc.

This will be briefly described here using the website of the Party Commission of the Uyghur Autonomous Region of Xinjiang as an example. The URL of this stored website is:

<https://wayback.archive-it.org/13340/20200508214909/http://uy.ts.cn/>

During the test crawls, many sub-categories and contributions, especially video contributions on this website were not saved, or the hyperlinks were not active in the saved versions. The daily edition of the respective newspaper is usually offered as a PDF on the website free of charge. However, the hyperlinks to the PDFs were initially not saved or not saved correctly. Therefore, the FID staff checked every single category of this website and reported the problems to the CRL. This solved most of the problems in the follow-up crawls.

Here is one example: “The Daily Newspaper of Xinjiang” (Chin. Xinjiang Ribao) is published in Chinese, Kazakh, Uyghur and Oirat (Western Mongolian).⁹ The FID concentrates on the last three languages. The content of the editions in the various languages differs to some extent. PDF editions in the above-mentioned languages can be found on the website of this newspaper. URL of the Uyghur version:

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

Link to the PDF version, which is mostly identical to the print version:

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

At the beginning of archiving there were problems with the storage of the PDF editions of this newspaper in the three languages mentioned above. The problem was reported to the CRL by the FID staff. The problem was solved.

These Newspaper websites are particularly valuable because they provide access to past issues, often retrospectively over years. Thus, an entire newspaper archive has been stored here. The newspaper websites are described in more detail below. More than 14,000 newspaper pages (as of 21.01.2021) have been stored as PDFs or images.

Example Daily Newspapers

Eight daily newspapers from Sinkiang in the autochthonous languages of the region have been stored in the Central Asian Web Archiving so far. In addition, there is the Uyghur weekly newspaper from Kazakhstan.

A total of over 14,000 newspaper pages have been saved as PDF or image files (as of 21.01.2022). For two newspapers, there are still problems with the storage of the PDFs, which so far could not be solved. Many newspaper websites contain archives of older issues going back many years. It happens that archives of older issues have been deleted from the current website but remain stored in the Central Asian Web Archiving. The following are examples of newspaper websites from Xinjiang.

Daily Newspapers of Xinjiang in Uyghur Language

The website of the Daily Newspaper of Xinjiang (Chinese: Xinjiang Ribao) in Uyghur, Kazakh and Oirat/Mongolian is stored in the Web Archive. The newspaper is issued daily with usually four pages and more pages for special issues. On the website not only the PDF of the printed version but also many contributions on politics, economy, society of Xinjiang and the whole of China are presented. The

⁹ Newspapers in a particular language are searched for in the GVK catalogue like this: Uyghur newspapers: aiw zeitung spr uig; Kazakh newspapers: aiw zeitung spr kaz; Mongolian newspapers: aiw newspaper spr mon.

PDF issues of the Uyghur version between 01.05.2019 and 05.04.2021 are available on the website <http://uy.xjdaily.com.cn> and in the web archive https://wayback.archive-it.org/13340/*/http://uyghur.xjdaily.com/). The PDFs of the Kazakh version between 15.10.2019 and 26.12.2021 are available in the web archive: <https://wayback.archive-it.org/13340/20211227050144/http://kazakh.xjdaily.com.cn/>. The Oirat/Mongolian version between 01.11.2020 and 26.12.2021 is also stored in the web archive: <https://wayback.archive-it.org/13340/20211227050133/http://mongol.xjdaily.com/>.

Since 05.04.2021 the Uyghur version of the newspaper website has been reconstructed and changed, so that the newspaper issues are now represented as embedded pictures which makes them difficult to crawl and save. The new homepage of the Uyghur Xinjiang Daily is: <http://uyghurwap.xjdaily.com/xjrb-wwb/2021/04/15/1.html>

On the old homepage before the site was reconstructed, the start page had four buttons to choose between the languages of Chinese, Uyghur, Kazakh, Oirat/Mongolian. The start page had eight sub-pages as of 18.12.2020: <https://wayback.archive-it.org/13340/20201218184343/http://uyghur.xjdaily.com/>.

The eight sub-pages of the old version are:

(Start page)

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

- Muhim xäwärlär (“important news”), URL: <https://wayback.archive-it.org/13340/20200527205908/http://uy.xjdaily.com.cn/xw/yw/index.shtml>
- Mämlikät Xäwärliri (“national news”), URL: <https://wayback.archive-it.org/13340/20200527210003/http://uy.xjdaily.com.cn/xw/gnxw/index.shtml>
- Şinjang xäwärliri (“news from the region of Xinjiang”), URL: <https://wayback.archive-it.org/13340/20200527210021/http://uy.xjdaily.com.cn/xw/xjhw/index.shtml>
- Jaylarda (“local news”), URL: <https://wayback.archive-it.org/13340/20200527210037/http://uy.xjdaily.com.cn/xw/dzxw/index.shtml>
- Obzor (“commentaries”), URL: <https://wayback.archive-it.org/13340/20200527210054/http://uy.xjdaily.com.cn/xx/pl/index.shtml>
- Öginiş (“learning”), URL: <https://wayback.archive-it.org/13340/20200527210120/http://uy.xjdaily.com.cn/xx/index.shtml>
- Sayahät (“tourism”), URL: <https://wayback.archive-it.org/13340/20200527210227/http://uy.xjdaily.com.cn/xw/lyxw/index.shtml>

- Sipirliq gezit (“digital newspaper” / PDF-newspaper), URL: <https://wayback.archive-it.org/13340/20200527210321-/http://uy.xjdaily.com.cn/shuzibao/index.html>

Table 1. Overview of saved newspaper websites from Xinjiang and Kazakhstan (18.01.2022)¹⁰

Nr.	Newspaper Web Site-Name/Link (Start Page)	Number of archived pages (PDF or pic.)	Subpages Example: Link	Number of archived sub-pages
1	“Šynjang gazeti” Xinjiang Daily in Kazakh language https://wayback.archive-it.org/13340/20210219185334/http://kazakh.xjdaily.com.cn/	2132 (PDF)		
2	“Šinjang geziti” Xinjiang Daily in Uyghur language https://wayback.archive-it.org/13340/20210212193834/http://uy.xjdaily.com.cn/	1920 (PDF)	https://wayback.archive-it.org/13340/20200807184218/http://uy.xjdaily.com.cn/xw/yw/index.shtml	8
3	“Šinjiyang-giyin ödör-in sonin” Xinjiang Daily in Oirat/Mongolian language https://wayback.archive-it.org/13340/20210212193834/http://mongol.xjdaily.com/	1080 (PDF)	https://wayback.archive-it.org/13340/20201016182747/http://mongol.xjdaily.com/126/index.html	14
4	“Altay gazeti” ¹¹ https://wayback.archive-it.org/13340/20200908145717/http://kazakh.altxw.com/news/altrb/index.shtml			

¹⁰ Because of a lack of funding, no further pages could be saved after December 31st, 2021.

¹¹ The “Daily News from Altay” is a Kazakh newspaper with eight pages for each day, except for weekends and holidays. The issues from February 2020 onward are available as an archive on the website in a different data format as embedded pictures. Archive.It has problems saving these pages. This news website has seven subpages. URL: <http://kazakh.altxw.com/>

5	“Ili geziti” ¹² (Uyghur) / “Ile gazetisi” (Kazakh) https://wayback.archive-it.org/13340/20210212182506/http://epaper.ylxw.com.cn/user/yrb.html			
6	“Tarbagatay gazetisi” ¹³ https://wayback.archive-it.org/13340/20210212182504/http://www.kztcxw.cc/category/58_1.html	8152 (PDF)	https://wayback.archive-it.org/13340/20200918184340/http://www.kztcxw.cc/category/18.html	
7	“Uyghur avazi” ¹⁴ (Weekly Uyghur newspaper from Almaty/Kazakhstan) https://wayback.archive-it.org/13340/*/http://uyguravazi.kazgazeta.kz/	528 (PDF)	https://wayback.archive-it.org/13340/20210212191248/http://uyguravazi.kazgazeta.kz/rubric/118/news	
8	“Xälq tori” ¹⁵ (People’s newspaper in Uyghur) https://wayback.archive-it.org/13340/20210212182642/http://uyghur.people.com.cn/	Only online no print or PDF	https://wayback.archive-it.org/13340/20210212195201/http://uyghur.people.com.cn/155988/309877/index.html	20

¹² The Daily News from Ili appear in Uyghur and Kazakh language on weekdays with four to eight pages. The PDFs are available on the news webpage, but Archive.It still has problems saving the PDF. The news webpage has 18 subpages. URL: <http://kazakh.ylxw.com.cn/>

¹³ The PDF online archive of the “Tarbagatay Daily” in Kazakh language for a period starting 01.11.2013 until 2021 with the exception of some issues from mid-September 2017 to mid-December 2017 have been saved in the Central Asian Web Archive successfully.

URL for the issue 01.11.2013: https://wayback.archive-it.org/13340/20210212191652/http://www.kztcxw.cc/category/58_60.html
https://wayback.archive-it.org/13340/20210212191652/http://www.kztcxw.cc/category/58_60.html

¹⁴ This weekly paper appears with eight pages which are available online as PDF. No archive of past issues is available online except for the issues saved in the Central Asian Web Archive.

¹⁵ Online version of the Chinese “People’s Daily Newspaper” (chin. Renmin Ribao) in Uyghur. The “People’s Newspaper” is a party newspaper of the Chinese Communist Party. The “People’s Newspaper” has no printed edition in the Uyghur language. This newspaper website has though different versions in different languages online only. Among them are the languages of seven national minorities of China. There are also versions in nine foreign languages. This online newspaper is available in the above-mentioned languages with different content.

9	“Xalıq torabı” ¹⁶ (People’s newspaper in Kazakh) https://wayback.archive-it.org/13340/20210212182517/http://kazakh.people.com.cn/ (in Kazakh Arabic script)	Only online no print or PDF	https://wayback.archive-it.org/13340/20210212182517/http://kazakh.people.com.cn/306608/index.html	21
---	--	--------------------------------	---	----

Cataloguing and Indexing Through Keywords and Translation

Individual Central Asian web pages were already stored in the Internet Archive’s Wayback Machine before 2019. Crawls for these pages were often many months apart. Hyperlinks were often not saved. The stored pages were hardly searchable unless the URL was known.

In order to enable a targeted thematic search for Central Asian web pages in the Internet Archive, about 500 Central Asian and especially Uyghur web pages with sub-pages have so far been catalogued in the ZDB (Zeitschriftendatenbank), indexed and a short description in English and German added. These catalogue records are accessible worldwide via the Common Union Catalogue (GVK). A short guide for searches in this catalogue is available under “How to search the Webarchive” : <https://fid-cassib.de/webarchive-central-asia/> Also: [Explanation_Webarchive_and_Search_26_3_21.pdf](https://fid-cassib.de/webarchive-central-asia/Explanation_Webarchive_and_Search_26_3_21.pdf) (fid-cassib.de)

Due to the complex structure of many websites, thematically deviating sub-pages have been catalogued separately.

Each catalogue record contains the following categories, among others, which can also be searched individually:

- publication date (first verified sighting)
- language code
- country code
- place of publication
- Title of the web page (parallel title, variant title and previous main title, if applicable. Original written record in the original script and transliteration of title)
- Authors/corporations/parties (original script and transliteration).
- Date(s) of crawls
- Subject headings (according to the rules of the ZDB, only six subject headings can be assigned to one title record).
- URL (current at the time of cataloguing, updated later if necessary)

¹⁶ Online version of the Chinese People’s Daily Newspaper (chin. Renmin Ribao) in Kazakh. The “People’s Newspaper” does not have a printed edition in Kazakh. This online newspaper in the Kazakh language is available in both Kazakh-Arabic script and Cyrillic script: URL: <https://web.archive.org/web/20220000000000/http://kazakh.people.com.cn/>.

- Link to the archived web page in the Internet Archive/Central Asian Web Archiving
- Access restriction, if applicable
- Content description (Because only six keywords are possible, the web page has an additional description in full sentences in English and German in field 4207).
- Font/script(s)

Due to the complex structure of many websites, thematically **deviating sub-pages** have been catalogued separately. Each catalogue record contains the following categories, among others, which can also be searched individually:

- publication date (proven since)
- language code
- country code
- place of publication
- Title of the web page (parallel title, variant title and previous main title, if applicable).
- Original written record and transliteration of title)
- Authors/corporations/parties (original written record and transliteration).
- transliteration)
- Mirror date
- Subject headings (according to the rules of the ZDB, only six subject headings can be assigned to a title record).
- URL (current at the time of cataloguing, updated later if necessary)
- Link to the archived web page in the Internet Archive/Central Asian Web Archiving
- Access restriction, if applicable
- Content description (Only six keywords are possible. Therefore, the web page is described again in English and German in bound language in field 4207)
- language)
- Font

Example of a Government Website from Sinkiang:

The recording of the entrance website of the Party Commission of Sinkiang initially looks as follows in the cataloguing system. The categorie 4207 contains also a description in English:

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

0501 Text\$btxt

- 0502 Computermedien\$bc
- 0503 Online-Ressource\$bcr
- 1100 2001\$n[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%%Tiyanshan tori / Šinjang uyğur aptonom rayonluq xälq hökümiti axbarat işxanisi; Näšir hoquqi Šinjang yengi taratqular märki-zigä täwä 4000 شىنجاڭ ئۇيغۇر ئاپتونوم رايونلۇق خەلق / تورى %T01\$UArab\$
 ھۆكۈمىتى ناخىبارات تىيانشان
 ئىشخانىسى: نەشىر ھوقۇقى شىنجاڭ 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: Tängritağ tori
 تەڭرىتاغ تورى :T01\$UArab%%Haupttitel 2001-2018\$ 4213
- 4221 In arabisch-ugurischer 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!Parteipolitik [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-goettingen.de/DB=1/LNG=DU/CLK?IKT=12&TRM=1689724374
 7800 666957975
 7900 25-03-20 17:01:01.000
 8001 %8g{cruuuuuuuuuuuu}
 8032 #1#
 8034 Bestand s. OPAC (Link anklicken)
 8597 h3609052775

A general problem in the indexing and cataloguing of web pages is the fact that most web pages are composed of several subcategories and, according to the cataloguing rules of the ZDB, each individual category is regarded as an independent title and requires its own title record. Therefore, in this example, if you search for this website in the Common Union Catalogue (GVK) under *Tiyanşan tori*, you will come across a total of 40 title records: One record for the main page and 39 more for sub-pages.

The Sub-Pages

1. muhim xäwär (Eng. “important news”), 2. Şinjang (Eng. “Xinjiang”), 3. mämlikäät (Eng. “Chinese inland”), 4. näzäriyā (Eng. “theory”), 5. ekran¹⁷: Män wä mening wätinim (Eng. screen: “I and my fatherland”), 5.1. Ekran: Män wä mening wätinim: Öginiş dawamlaşmaқта (Eng. screen: “I and my Fatherland: Learning progresses”), 5.2. Ekran: Män wä mening wätinim: Yengi дәwr, yengi nätijä, yengi sähipä (Eng. screen: “I and my Fatherland: New era, new achievement, new page”), 5. 3. Ekran: Män wä mening wätinim: Sältänätlik 70 yil küräşçän yengi дәwr (Eng. screen: “Me and my fatherland: The great 70 years – a fighting new era”), 5. 4. Ekran: Män wä mening wätinim: Namratliqtin qutulduruş ötkiligä hujum qiliş jengidä qätiy gälibä qilayli (Engl. screen: “Me and my fatherland: Let us fight in the storm, struggle against the hurdle of poverty and surely win”), 5. 5. Ekran: Män wä mening wätinim: Määdäniyättä Şinjangğa näp yätküzüş (Eng. screen: “Me and my fatherland: To win in the field of culture of the Xinjiang region”), 5. 6. Ekran: Män wä mening wätinim: Millätlär ittapaqlışip bir a’ilä kişiliridäk boluş (Eng. screen: “I and my fatherland: Nationalities become members like members of one united family”), 5. Ekran: Män

¹⁷ Ekran, a Russian foreign word in Uyghur, means screen and usually refers to videos. If you click on the “Ekran” tab on the top, a page opens with numerous video categories, all of which are listed here as examples. “Screen” (Ekran) can often be better translated as “media library”.

wä mening wätinim: Yaşliq küräş qiliş üçündür (Screen: “I and my fatherland: Man’s youth is for struggle”), 5. 8. ekran: Män wä mening wätinim: Jungxua änäniwi mädnäniytini äwj alduruş (Screen: “I and my fatherland: Spreading traditional Zhonghua¹⁸ civilisation in Xinjiang), 6. Räsimplik yäşmä (screen: “Narrating through Pictures”), 7. Äsliy mäqsätini untumay, burçni ästä çing saqlaş (screen: “Keeping the Original Objectives in Mind and Taking the Mission to Heart”), 8. Kündä bir altun jümlä (Eng. “Learn a sentence worth its weight in gold every day”), 9. Wätänpärwärlük hesyati, küräş qilgüçilar (Eng. “Patriotism and fighting”), 10. Tiyanşan tori mäxsus sähipisi: Iradidin sepil hasil qilip, yuqumning aldini eliş – tizginläş, tosüp zärbä beriş jengidä gälilibä qilayli (Engl. Special Forum of Tiyanşan: “Let us build a firm defensive wall from our will and let us thereby achieve victory in the field of struggle in the prevention of infection, in the control measures against infection and in the attack against infection”), 11. Sältänätlik 70 yil küräşçan yengi дәwr (Eng. “The Magnificent 70 Years: Combative New Era”), 12. Şinjangni qalaymiqan qiliş xam xiyalida bolğan härqandaq suyiqäst härgiz mäqsitigä yetälmäydu; Şinjangdiki här millät xälq qätiy qoşulmaydu (Eng. “Any conspiracy with empty fantasies to throw the Xinjiang area into chaos must not be allowed to reach its goal under any circumstances – All nationalities in Xinjiang are determined not to allow this.”), 13. Şinjang yaxşı jay (Eng. “Xinjiang is a good place”), 14. Millätning güllinişi üçün: Qährimanlar, qurbanlar şäjärisi (Eng. “For the Revival of the Nation: The Biographies and Life Stories of Heroes and Sacrifice Makers”), 15. Şinjangdiki bayqaş (Eng. “Discoveries in Xinjiang”), 16. Ätimiz teximu güzäl: Käspi maharät tälim – tärbiyä: Ümidkä tolğan güzäl turmuş (Eng. “Our future will be even more beautiful: education and training in professional skills: the beautiful life full of hope”), 17. Güzäl yurt – makanimizni yaxşı qoğdayli (Eng. “Let us defend our beautiful homeland well”), 18. Namratliqtin qutulduş ötkiligä hujum qiliş (Eng. “The struggle against the difficulties in the fight against poverty”), 19. Älräyini biliş, älgä näp yätküzüş, älni mayil qiliş (Eng. “Be aware of the people’s feelings and opinions, bring profit to the people, win the people’s hearts for the government”), 20. Millätlär ittapaqlişip bir a’ilä kişilridäk boluş (Engl. “Nationalities become members like family members of a family by uniting”), 21. Bingtuän (Engl. “(Xinjiang Production and Construction Corps), 22. Şinjangğa nişanlıq yardäm beriş (Helping the Xinjiang region in a targeted way), 23. Ürümçi (Urumchi), 24. Mädnäniyät – sayahät (Culture and tourism), 25. Jaylarda (Reports and news from the districts and cities in Xinjiang), 26. Pän – ma’arip (Engl. “Education in Xinjiang”), 27. Yeza igiligi (Engl. “Agriculture in Xinjiang”), 28. Täntärbiyä (Engl. “Sport”), 29. Turmuş luğiti (Engl. “Living World”), 30. Şinjang näzäriyä tori (Engl. “Special Forum of Xinjiang for Political Theory”).

The Uyghur titles of the above sub-pages can be searched individually in the GVK.

¹⁸ Zhonghua: Chinese for “middle culture”, i.e. Han Chinese culture. The term often stands for China per se.

When cataloguing, it is often necessary to add: even if the country code, date of publication and corporate body(ies) of this webpage can be determined from the information found on this webpage, field 4207 and the fields between 5100 and 5105 must still be filled in later, after having made an overview of the content of all categories of this webpage. In the context of a ZDB record, field 4207 is information on the content of electronic resources, while the fields between 5100 and 5105 are made up of keywords and thus also make the content of the medium in question accessible by means of keywording.

In many cases, there is no information on the respective web pages about whether the main title has changed over the years and when this might have happened. But title changes and old titles must also be researched and entered. These questions are of enormous importance for building a high-quality web archive and for good formal and subject indexing of web pages.

In order to solve the aforementioned problem of the former main titles of the Party Commission's website, extensive research in the Internet Archive <https://archive.org/web/> was necessary. Our FID staff looked there at the issues of the website in question from earlier years and tried to find answers to questions. This means more time and effort to merge old and new versions of a website.

On this party website there are videos on many sub-pages as described above. Most of them were saved during the mirroring by Archive.It. When called up, it often takes a long time for the videos to load.

Here are two examples: The first video is called: "Baş şujining bir häptisi" (Eng: "A week in the life of the General Secretary"):

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

The second example is the following video: "Xotän wilayiti keriyä nahiyisidiki täsirlik hekayä" (Eng: "An impressive story in Keriyä County of Hotan District"):

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

In the case of many Uyghur exile websites, it is difficult to determine the country code, date of publication, place of publication, as these details are often missing from such websites for various reasons.

Questions and problems about content or technical aspects often remain unsolved for the time being.

Literature

- Äysa, Äziz Älkun (Azia Isa Elkun) (2011): “Uyğur tor bätçilikigä omumiy nazar: Uyğurlarda internetniñ qollinişi – ärkin pikir bayan qilişniñ wastisi” [A General overview of Uyghur websites: The use of the Internet by the Uyghurs as a means of freedom of expression] (50 pages), this essay is available in Uyghur online at the following link
http://www.uyghurpen.org/uy/Uyghur_torbetchilikige_omumiy_nezer.pdf
- Beinert, Tobias/Hagenah, Ulrich/Kugler, Anna (2014): “Once upon a time there was a website. Kooperative Webarchivierung in der Praxis”, in: *Das offene Bibliotheksjournal*, H.1, 291–304, this article is available at the following link:
<https://www.o-bib.de/article/view/2014H1S291-304>
- Schumacher, Edgar (2016): “Webarchivierung an der SUB Hamburg: Ziele, Erfahrungen, Perspektiven” (20 pages), this article is available at the following link: [https://www.babs-muenchen.de/content/veranstaltung/2016-03-10/2016_03_10_schumacher.pdf10/2016_03_10_schumacher.pdf](https://www.babs-muenchen.de/content/veranstaltung/2016-03https://www.babs-muenchen.de/content/veranstaltung/2016-03-10/2016_03_10_schumacher.pdf10/2016_03_10_schumacher.pdf)
- Rauber, Andreas/Liegmann, Hans (2010): “Web archiving for the long-term preservation of internet documents”, in: *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, ch.17: 88-Kap.17: 103, Glückstadt: Verlag Werner Hülsbusch.