



New ways of content discovery: the CrossAsia Newspaper Research Companion

Anna Stryjewska

CrossAsia

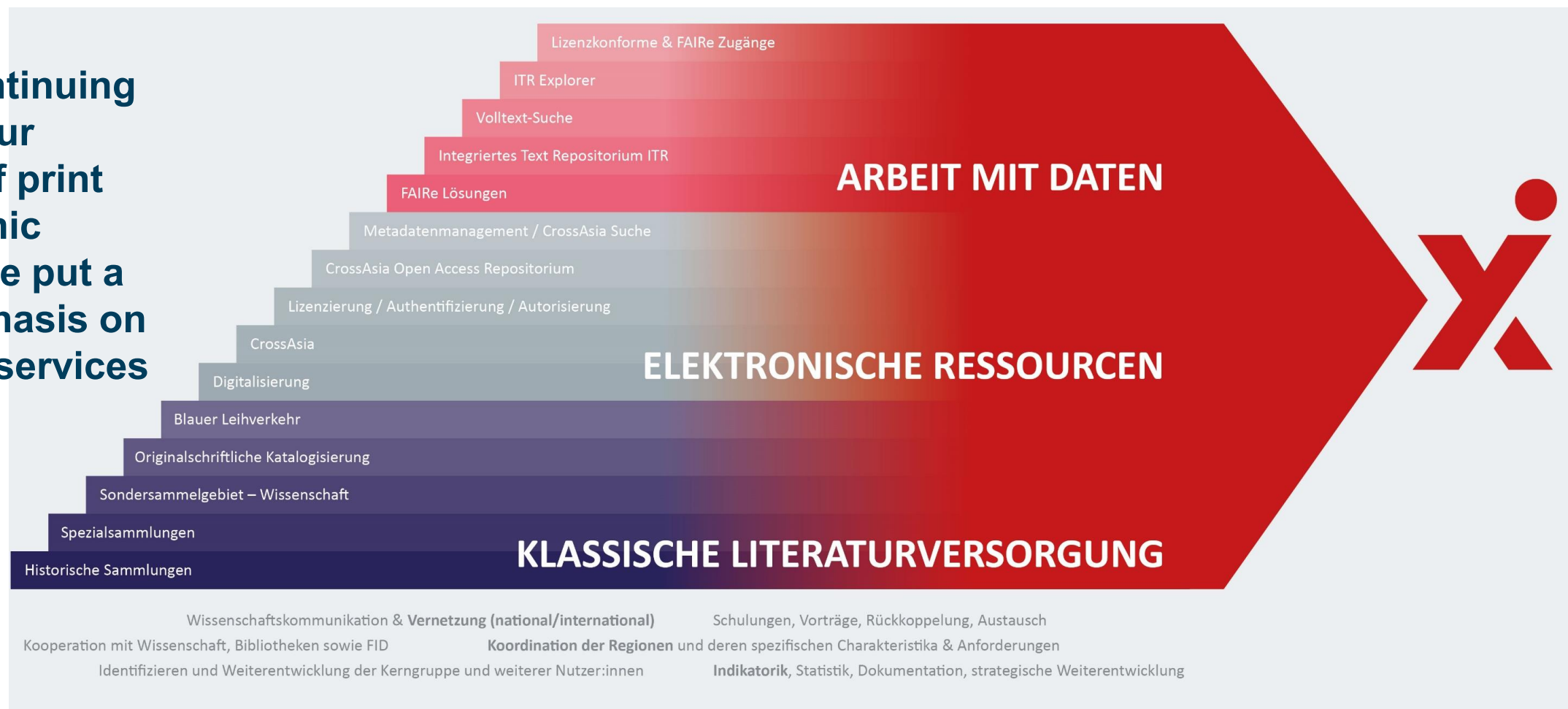
Staatsbibliothek zu Berlin

anna.stryjewska@sbb.spk-berlin.de

EASL 2026, CATS Heidelberg

Our mission at CrossAsia

Besides continuing to expand our collection of print and electronic materials, we put a strong emphasis on data-based services



Our mission at CrossAsia

- Integrating our own and external data from licensed databases into a single infrastructure, the Integrated Text Repository (80 million+ digital objects)
- Rather than just archiving data, our goal is to develop discovery tools which allow readers to explore it in full depth

digitized data →

intelligent processing →

discovery of **relevant** material

Library 4.0

NLP

AI

Artificial Intelligence

NLP

BIG DATA

BIG DATA

Img source: www.chatgptlibrarian.com

CrossAsia Integrated Text Repository

Full-text search

Current Selection

remove all

(x) Cotton

Search

Collection

- ☐ North China Daily News (1912)
- ☐ SBB digital : Western language Asia collection (1142)
- ☐ The North China Herald Online (1919)
- ☐ Adam Matthew - Foreign Office Files China & Japan (1912)
- ☐ Japan Chronicle Online (1919)
- ☐ Mobilizing East Asia, 1931-1954 (1917)
- ☐ Early Twentieth Century Chinese Books (1912-1949) (1912)
- ☐ Area Studies Japan, China, and Southeast Asia (1919)
- ☐ North China Standard Online (1917)
- ☐ China and the Modern World: Records of the Maritime Customs Service of China (1854-1949) (1919)

Type of Object

- ☐ Page (1919)
- ☐ Article (1919)
- ☐ Book (1917)

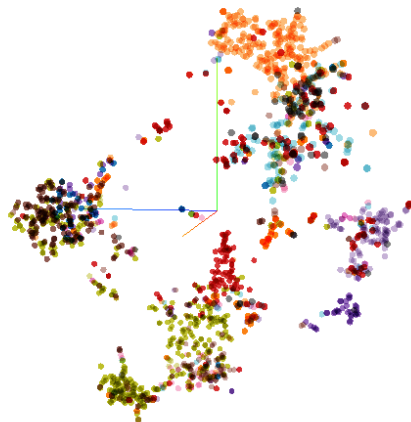
Title

- ☐ Imperial Maritime Customs 134 (1919)
- ☐ Dai-Nihon-gaikoku-beki-tenpō (1912)
- ☐ Commercial reports by Her Majesty's consuls (1919)
- ☐ Imperial Maritime Customs (1912)
- ☐ The Maritime Customs of China

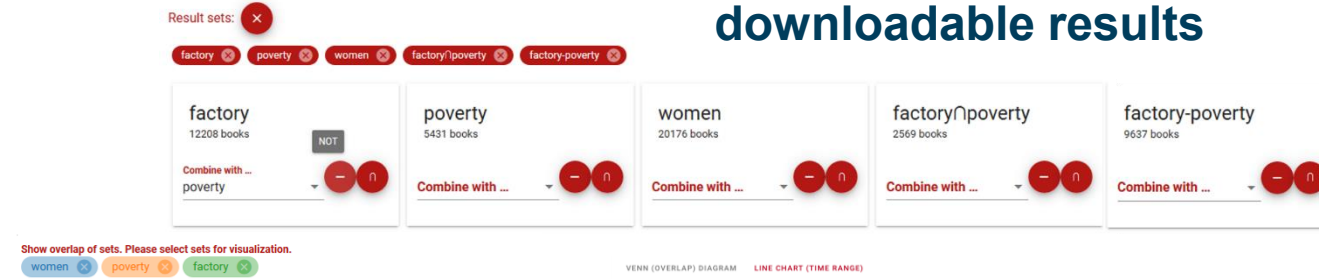
... (Continued) Ching Roong Co
Sing Cotton s &w. Co...

citation:
collection:
CrossAsia licence:

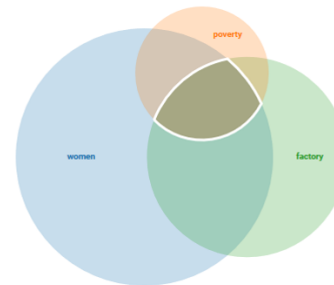
N-gram data sets



ITR Explorer: data visualization, downloadable results

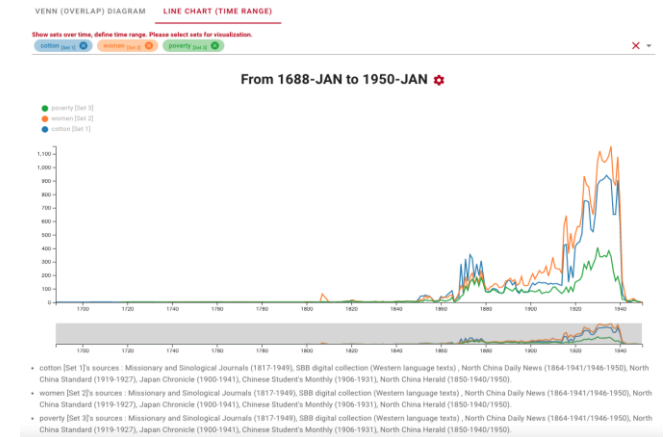


(Set "factory" ∩ Set "poverty" ∩ Set "women") : 2213 books
that is
10 % of the set "women"
40 % of the set "poverty"
18 % of the set "factory"



Titles in selection ((Set "factory" ∩ Set "poverty" ∩ Set "women") : 2213 books)

Title	Sets	Date	Creator	Subject	CrossAsia
The Rev George Smith DD Bishop of Victoria Hong Kong	● ● ●	n.d.	Headla...	Hong K...	↗
Bookshelf The English-language Press Networks of East Asia 1918-1945	● ● ●	2010-01-01			↗
澳門現代化進程與城市規劃	● ● ●	2007	楊允中	社會科學 -...	↗



Newspaper Research Companion

- Newest **Open Access** service within the CrossAsia ITR
- The problem: newspaper data scattered across archives in different languages
- Our answer: An AI Platform for cross-lingual newspaper archive research

In collaboration with



Sponsored by



Der Beauftragte der Bundesregierung
für Kultur und Medien

Newspaper Research Companion

- **Diverse Newspaper Archive Corpus**

Over 70M articles from 253 newspapers in German, Chinese, English, Russian spanning 1785–2026 (incl. in-house and licensed material)

- **Key titles:** *Berliner Tageblatt, Ta Kung Pao, Times Digital Archive, People's Daily, Pravda, South China Morning Post*

- **Semantic Search**

Understands intent behind the query, not just keywords (but **not** a chatbot!)

- **Cross-Lingual Discovery**

- **Results strictly based on the data corpus (no hallucinations)**

Let's try it out!

www.crossasia.org/nrc

Why Early Access?

- First iteration of service with continuous improvements ahead
- Licensing agreements still being finalized
- Collection will expand significantly beyond current 70M articles; more languages will be added
- Design and features refined based on **user feedback**
- OCR developments
- We are still developing the best method to query the NRC



Thank you!

Share Your Feedback

Try it out: [the NRC homepage](#)

Share your thoughts: x-asia@sbb.spk-berlin.de

Share With Your Readers

For researchers already working with newspaper data
or those who might discover its relevance (**1785–2026**)