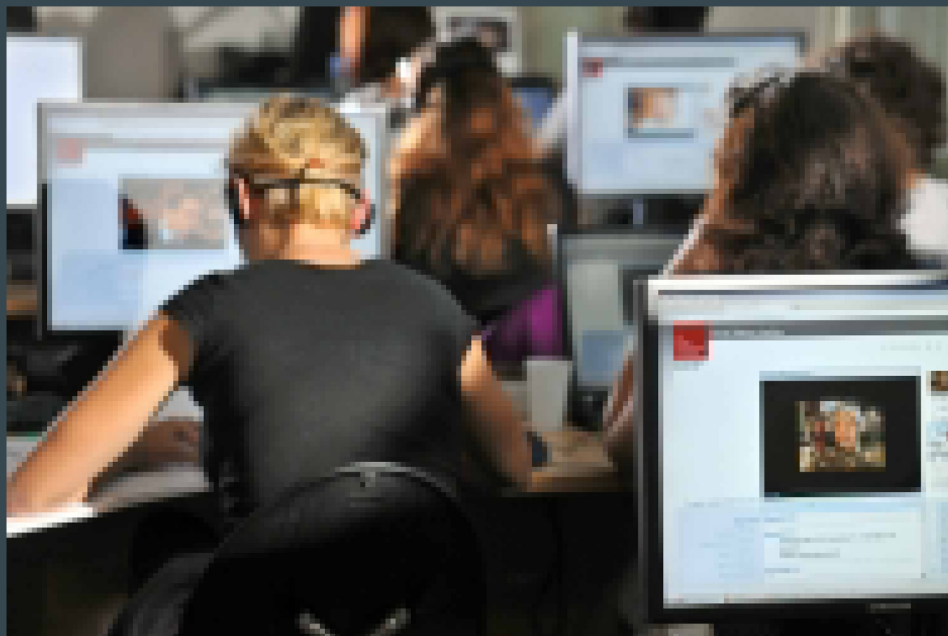




Data Analytics for Social Sciences and Humanities

Malach Centre for Visual History



Karin Hofmeisterová
Jiří Kocián

malach@ufal.mff.cuni.cz



Multilingual Access to Large Spoken Archives

- Goal of the project (2001–2006): to use automatic speech recognition (ASR) and information retrieval (IR) techniques to provide improved access to the large multilingual spoken archives created by the USC Shoah Foundation
- <https://malach.umiacs.umd.edu/>
- Establishing Malach Centre for Visual History in 2009/2010 as a follow-up project at CUNI





What do we do?

- Providing access to large archives of audiovisual oral history testimonies of genocide survivors (mostly Holocaust)

Activities

Research support:

- providing on location access to data
- end user-interface help and troubleshooting
- sampling and search service
- software testing

Education and training:

- seminars and activities for students (elementary, high schools, universities)
- seminars for researchers
- methodical seminars for teachers

Public Access:

- survivors and relatives
- public events, consultancy



Archives at CVHM



- **VHA USC Shoah Foundation**
 - Access to 54 975 interviews in 42 languages recorded in 1990s
 - Focus on Holocaust (1939-1945)
 - VHA continuously enlarged by new collections related to other genocides
 - catalogued, indexed
- **Fortunoff Video Archive for Holocaust Testimonies**
 - Available from 2018
 - 4.400 interviews, 163 in Slovak and 17 in Czech
 - catalogued, keyword search
- **AJR Refugee Voices**
 - 150 interviews with Jewish refugees in Great Britain
- **The Jewish Holocaust Center in Melbourne**
 - 15 interviews with Jewish survivors born in Czechoslovakia

Each interview is a chronological life story (including pre-genocide and post-genocide events)

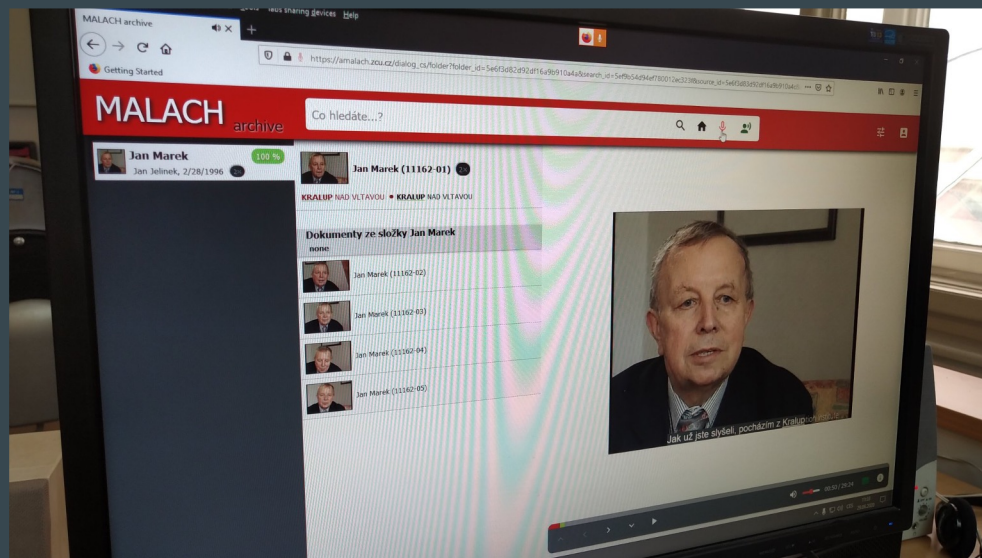
Approximately 2–3 hours long

Relevant for multiple disciplines (not only history but also sociology, psychology, linguistics...)

AMALACH



- Phonetic, fulltext search tool (developed by University of West Bohemia)
- Search within VHA collections of testimonies with the Holocaust survivors in Czech, Slovak and English languages



New projects and goals

- Informing about less known mass violence in Europe through oral history
- Broadening collections available at CVHM
- Preserving testimonies from oblivion
- Unique collection Zaveštanje – child survivors of genocide in the Independent State of Croatia



MALACH User Interface

- Developing user interface, which integrates search within all available archives
- Prepared to integrate new collections
- Aim to minimize deficiencies of particular search engines and methodologies
- Integrating open access archives, partially access archives and archives accessible only in CVHM



The screenshot displays the MALACH user interface. At the top, there is a navigation bar with logos for 'cvh mach', 'LINDAT', 'UFA', and 'FACULTY OF MATHEMATICS AND PHYSICS Charles University'. A search bar is located on the right side of the navigation bar. Below the navigation bar, there is a menu with links to 'USC SF Visual History Archive', 'Yale Fortunoff Video Archive', 'USHMM', 'Centropa', 'Refugee voices', 'JHC Melbourne', 'Malach Search', and 'Resources'. The main content area features a quote: "USC Shoah Foundation's Visual History Archive is the world's largest collection providing access to approximately 55.000 interviews of the Holocaust and other genocides survivors through a well developed system of metadata and keywords search". Below the quote, there are three logos: 'USC Shoah Foundation' with the text 'VHA Website' underneath, 'AMALACH' with a red square icon containing a white waveform, and 'USC Shoah Foundation IWITNESS' with the text 'IWITNESS' underneath.

Activity 1 - MALACH interface (10 min)

- Try to find in the available databases the interviews associated with your birth country/town
 - Were you able to reach a result with all of the databases?
 - Which archive gave you the largest set of relevant results?
 - Were you able to recognize differences between search technologies?

Activity 2 – VHA (10 min)

- VHA – continue as a guest
- Task 1 - Find testimonies associated with your countries of origin.
 - What is the role of the respective country in the story-line of a selected testimony? (is it a birth country, transit country, place of the interview?)
 - How did you find out?

Activity 3 – AMALACH (15 min)

- Task 1 - Find interviews (either in Czech or English), in which your birth-country is mentioned.
 - How does the search differ from VHA?
 - Compare the relevance of results.
- Task 2 - Find discrepancies between the speech and automatically generated subtitles. What kind of words it concerns? Can these discrepancies negatively affect the understanding of the meaning? How do they affect the search and its functionality?

Reflection and discussion

- Share your impressions of working with various search tools
- What are the limits of various search tools?
- Any suggestions for improvement? :)