Matthias Withopf

Software Engineering Consulting

April, 2005

The EYDES archive: From Tape Reels to the Internet

The sound archive of The Language and Culture Archive of Ashkenazic Jewry (LCAAJ) consists of nearly 6000 hours of spoken Yiddish. Between 1959 and 1972 interviews were conducted using a questionnaire of 3840 questions. Yiddish-speaking people from more than 600 different locations in Europe were interviewed, and the interviews were recorded on 2552 reels of tape. To make this treasure widely available, one important part of the EYDES project is to bring the sound archive to the Internet.

Intentions

- To access the original sound: from about 6000 hours in total.
- To access the locations: from 600 in total.
- To access the word index of all available transcripts.

You can look up a word in the word index and find and listen to the occurrences in the original sound.

We have 2 word indexes: one using Yiddish with Hebrew characters and one using the YIVO transliteration with Latin characters.

To link all data and navigate in the huge amount of sound so that not only experts and researchers can use the archive. Our goal is that everyone can access the sound archive and find useful and interesting information, even if he or she is not an expert and cannot read Hebrew letters.

Additionally, to just viewing the original information, we adapted a software application to apply statistical methods to the data (Lara). The results can be related to a map of Europe. The goal is not just to show the data, but to explore it and show how it correlates and find out how the single pieces work in context.

Extent of Data

2552 reels of tape (an analog copy of the data) were digitized and transferred to Audio CDs (a digital copy). This work was done at Columbia University, New York. Both the reels and the Audio CDs are located at Columbia.

We transferred the Audio CDs to a single computer hard disk using the MP3 audio compression. This hard disk currently holds about 60 gigabytes of information. So we compressed all the shelf space used for the reels to a single data medium.

We also use DVD-R (Double Layer) as backup media. All data fits on only 8 DVD-Rs.

For data safety reasons, we store one copy of the sound archive in a bank safe.

Security of Data

At the moment, we have two active servers (the main accessible one on website www.eydes.de and one mirror server),

three backups of the data pool on external hard disks,

several backups on DVD-R

With all backup copies of the sound and the transcripts kept in different locations, the storage of the data is considered safe.

Accelerating the process

Right now (April, 2005), we have the sound of about 479 of 603 localities converted, which is more than two thirds of the complete archive. Because it requires a lot of manual work to read in the Audio CDs, we have developed an automated process using a CD autoloader with a robotics arm to insert and eject the CD media.

Now we can insert a stack of 100 Audio CDs, press a button and come back when everything is done. So we hope to complete the remaining conversion work very soon.

Try it for yourself, at home or in any Internet Cafe

www.eydes.de