

Browsing and Searching the Spoken Words of Buchenwald Survivors

Roeland Ordelman Willemijn Heeren Arjan van Hessen Djoerd Hiemstra
Hendri Hondorp Marijn Huijbregts Franciska de Jong Thijs Verschoor

Human Media Interaction
University of Twente, Enschede, The Netherlands

1 The Buchenwald demonstrator

The ‘Buchenwald’ project is the successor of the ‘Radio Oranje’ project that aimed at the transformation of a set of World War II related mono-media documents –speeches of the Dutch Queen Wilhelmina, textual transcripts of the speeches, and a database of WWII related photographs– to an attractive online multimedia presentation of the Queen’s speeches with keyword search functionality [6, 3]. The ‘Buchenwald’ project links up and extends the ‘Radio Oranje’ approach. The goal in the project was to develop a Dutch multimedia information portal on World War II concentration camp Buchenwald¹. The portal holds both textual information sources and a video collection of testimonies from 38 Dutch camp survivors with durations between a half and two and a half hours. For each interview, an elaborate description, a speaker profile and a short summary are available.

The first phase of the project was dedicated to the development of an online browse and search application for the disclosure of the testimonies. In addition to the traditional way of supporting access via search in descriptive metadata at the level of an entire interview, automatic analysis of the spoken content using speech and language technology also provides access to the video collection at the level of words and fragments. Research in this phase was dedicated to the automatic annotation of the interviews using speech recognition technology [5] and combining manual metadata per interview with the within-interview automatic annotations for retrieval of both entire interviews and interview fragments, given a user’s query [4]. Moreover, having such an application running in the public domain allows us to investigate other aspects of user behavior next to those investigated in controlled laboratory experiments.

The second stage of the project aims at (i) the optimization of the automatic annotation procedure [1], (ii) interrelating all available multimedia resources, such as written summaries to exact video locations or locations to maps or floor plans [2], and connected to this, (iii) further development of a user interface that allows for the presentation of this information given the various user needs.

While survivors of World War II can still personally tell their stories, interview projects are collecting their memories for generations to come. Such interview collections form an increasingly important addition to history documented in written form or in the form of artifacts. Whereas social scientists and historians typically annotate interviews by making elaborate summaries or sometimes even full transcripts, by assigning keywords from thesauri and by establishing speaker profiles, catalogs based on these manually generated metadata do not often contain links into video documents. That is, they do not support retrieval of video fragments in response to users’ search queries; results are typically entire videos - that may be hours long - or (parts of) the transcripts.

The interview browse and search application of the ‘Buchenwald’ portal shows a multimedia search application based on both the conventional, manual metadata as well as automatic speech recognition output. It is part of a website on Buchenwald maintained by the Netherlands Institute for War Documentation (NIOD) that gives its user a complete picture of the camp then and now by presenting written articles, photos and the interview collection.

¹Buchenwald: <http://www.buchenwald.nl>

After the audio tracks had been separated from the video documents, the audio was processed by the open source speech recognition toolkit SHoUT² developed at the University of Twente, resulting in coherent speaker segments and a time-stamped transcript for indexing. For retrieval, the open source XML search system PF/Tijah is being used³.

The user interface supports browsing and search in the collection. To start browsing the collection, a user can request a list of all available videos. Each result contains links to the short summary, the speaker's profile, the elaborate description and the video document (Figure 1). To search the collection, a standard text search field is provided. If results are found, they are listed in the same format as the browse list, with the alteration that two types of results are available: interview results and fragment results.



Figure 1: Screen shots of the result list, showing the short summary, the speaker's profile and the video browser

Interview results imply hits in the textual, manual metadata, and fragment results imply hits in the speech recognition output. In the former case, the terms matching the user's query are highlighted in color in the textual metadata. In the latter, the video link directs the user to the exact speaker segment that contains the hit.

References

- [1] F.M.G. de Jong, D. Oard, R.J.F. Ordelman, and S. Raaijmakers. Searching spontaneous conversational speech. *SIGIR Forum*, 41(2):104–108, 2007. ISSN=0163-5840.
- [2] F.M.G. de Jong, R.J.F. Ordelman, and M.A.H. Huijbregts. Automated speech and audio analysis for semantic access to multimedia. In *Proceedings of Semantic and Digital Media Technologies, SAMT 2006*, volume 4306 of *Lecture Notes in Computer Science*, pages 226–240, Berlin, 2006. Springer Verlag. ISBN=3-540-49335-2.
- [3] W.F.L. Heeren, L.B. van der Werff, R.J.F. Ordelman, A.J. van Hessen, and F.M.G. de Jong. Radio oranje: Searching the queen's speech(es). In C.L.A. Clarke, N. Fuhr, N. Kando, W. Kraaij, and A. de Vries, editors, *Proceedings of the 30th ACM SIGIR*, pages 903–903, New York, 2007. ACM.
- [4] Djoerd Hiemstra, Roeland Ordelman, Robin Aly, Laurens van der Werff, and Franciska de Jong. Speech retrieval experiments using xml information retrieval. In *Proceedings of the Cross-language Evaluation Forum (CLEF)*, 2007.
- [5] M.A.H. Huijbregts, R.J.F. Ordelman, and F.M.G. de Jong. Annotation of heterogeneous multimedia content using automatic speech recognition. In *Proceedings of SAMT 2007*, volume 4816 of *Lecture Notes in Computer Science*, pages 78–90, Berlin, 2007. Springer Verlag.
- [6] R.J.F. Ordelman, F.M.G. de Jong, and W.F.L. Heeren. Exploration of audiovisual heritage using audio indexing technology. In L. Bordini, A. Krueger, and M. Zancanaro, editors, *Proceedings of the first workshop on intelligent technologies for cultural heritage exploitation*, pages 36–39, Trento, 2006. Universit di Trento. ISBN=not assigned.

²SHoUT ASR toolkit: <http://wwwhome.cs.utwente.nl/~huijbreg/shout/>

³PF/Tijah: <http://dbappl.cs.utwente.nl/pftijah/Main/HomePage>