Multimedia database management for annotators of the metadata content

Matúš Pleva, Jozef Juhár, Anton Čižmár

Faculty of Electrical Engineering and Informatics, Technical University of Košice

***Abstract* — Management of the multimedia databases of annotated multimedia content which are permanently growing is not easy. There are several difficulties when we want to permanently add some new content every day, and manage the annotation without problems. Also the process is not finished by annotating the multimedia. The annotators make always errors, and it is not possible to have the annotation perfect. But when we need the database for training the automatic speech recognition models, we need as good annotation as it is possible. Very good way is to build a multimedia interface for collecting annotation errors reports presenting multimedia using captions. Every user then can send a collection of errors he finds during the watching the captioned video or listening to captioned audio record. Then the group of authorized annotators could repair the original annotation using also a unified interface.**

***Index Terms* — management, metadata, multimedia, speech recognition, unified interface.**

I. Introduction

THESE days, there is a big demand for spoken dialog services, were a computer needs to understand the human and also synthesize the human speech. The next big challenge is to build an automatic metadata extractor for audiovisual content, for example for monitoring of the broadcast news (TV or radio). The automatic continuous speech recognition module is the essential part of the spoken dialog system [1]. The speech recognition based on Hidden Markov Models is built on the acoustic and language models, but they need a lot of training material. The training material consists firstly from annotated audio materials (for example SpeechDat [2]). For metadata extraction, they are also: speaker name, gender, acoustic condition, etc. annotated too. The video material is then usable for identifying of the speaker name, and also for better presentation of the annotated content to the user [3].

This paper describes the projection of the multimedia management interface, which is very useful and could lead to better annotation accuracy, which is essential for better acoustic models training. Annotating of the huge multimedia material using plenty of annotators needs automatic database management also because every annotator makes mistakes and the correction process is always time consuming. Also listening to the content of the audio file in which content the annotator is not interested leads to lack of concentration. The captioned video file using the data from the annotators is more interesting and also when the user of the database is interested in the content, he can easily boost the concentration.

The management interface for the annotators gives them a very good possibility to manage all annotations and corrections as simple as it is possible. The interface offers the new errors in the finalized annotations, which was reported from users of the presentation interface, and also new materials recorded from the broadcasters. When some annotator starts the correction or annotation of some material, the multimedia content is blocked to the rest of the annotators.

This management interface allows also logging of all the changes in the final database, so the administrator of the database could see who and when makes a new annotations or corrections of the metadata content.

II. The presentation interface

The presentation interface is responsible for authentication of the user and enabling him to see captioned video or audio files in the webpage interface embedded media player from the actual annotated database, and send an error report to the annotators including the time stamp of the position in the multimedia content (inserted automatically by the user-side scripting). Then the information is available to all annotators including the original transcription and all the multimedia content.

For captioning the multimedia files, the transcriptions from the database are converted online from STM (SDR99 Speech Recognition Task Time Segment Time Marked File) [4] to SAMI format, developed by Microsoft [5]. The free transcription tool Transcriber [6], which is used for annotation of the database content uses his own XML format TRS. But there is a possibility to convert or export to STM format, which is widely used in our laboratory when using the data for language or acoustic model training.

The biggest advantage of SAMI captions is that except spoken words, there is a possibility to present a speaker name in the captions (the name is presented by another color and font above the transcription called “SourceID line” [7]), and also other metadata like gender, emotion, etc. It means that all data that needs to be verified by the annotators are presented at a time and in a very intelligible way as we can see on the Fig.1.

Fig. 1. Presenting the annotated video using embedded windows media player in the web page using SAMI captioning functionality will help the annotators to find errors in the transcriptions and also in the metadata content as “name”, “gender”, etc.

For browsing the database, the open source code of the WebFileBrowser [8] was used and modified. The whole communication and reports was translated to the Slovak language as we can see on the Fig 2. The structure of the database was also clarified by directory structure composition, when different years of capture, different broadcasters and different quality subdirectories should help the annotators to find the appropriate content. The presentation interface also brings only one filename for video, audio, and annotation files, which differs only in the filename extension, so the number of the presented files is fewer.

The authorization to the presentation interface is made trough the MySQL database [9], which is connected with the management interface and gives a better control over the annotations to the administrator, who could also extract from the user logs the information and generate the statistics about who made how much new annotations, and how much corrections then they needs. The quality of the annotations and annotators is the main reason to do this statistics, because the quality is one of the essential factors for the acoustic models training.

III. Error Management Interface

One of the reasons for building the presentation interface is a possibility to send an error report about the presented captioned multimedia material to the error management database. The presentation interface sends also the time stamp of the report in the presented media. These reports are collected and all annotators authenticated for database correction should be able to see the error reports and easily find the appropriate files needed for the correction.

The developed web interface brings also the information about the new media (video or only audio) which was recently uploaded to the database by the recording engine. The proper annotator could also start a new annotation and after the finalization, upload the transcription to the database, and the management interface then automatically converts all files in the standard database structure and includes the content in the annotated part of the database. Then the content is available as a captioned video or audio using the presentation interface immediately.

To make the annotation and correction process as easy as possible, the management interface was developed and tested. As we can see on the Fig. 2, the number of the new multimedia content without any metadata is presented on the main screen of the authenticated annotators, besides the number of the new error reports, which was received from the presentation interface.

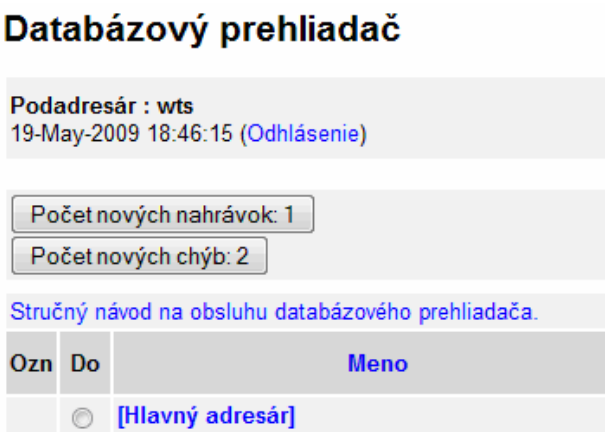

Fig. 2. Presenting the number of the new non-annotated content on the web page and also the number of the new error reports, which are not corrected until now.

After some annotator starts to annotate a new multimedia content or correcting the error found in the error report, he can easily download all appropriate files and they are not accessible anymore for the other users. When the annotator executes the buttons with this information, he can see that previously selected content is presented to the annotator as “in the progress” as we can see on the Fig 3a). And on the next Fig 3b) the annotator could see the handled error reports, where we can see that the annotation files, that some annotator download, because he start the annotation or correction of the annotation, are not accessible to the others.

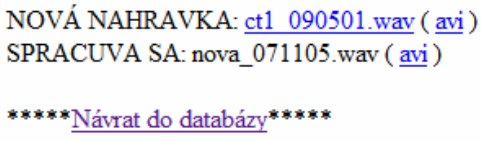

Fig. 3a. Presenting the list of the new non-annotated content on the web page and also the content of the database, which annotation is in progress.

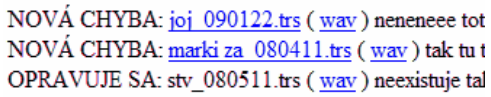

Fig. 3b. Presenting the list of the new error reports on the web page and also the errors, which are handled by the annotator and the correction, was not uploaded until now.

The second part of the error management interface offers the upload of the new or corrected metadata files in the database, and the server side PHP scripts [10] generates automatically captions in the SAMI format and move all appropriate content in the presentation database, which is also a source for the acoustic and language models training scripts.

The number of all unfinished material is clearly presented on the buttons on the main menu of the interface depicted on the Fig 4. below.

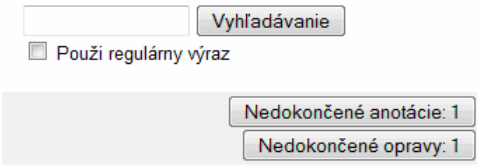

Fig. 4. Presenting the number of the unfinished annotations of the annotator: the unfinished new annotations on the top and the unfinished corrections below.

Then the new or corrected annotations could be uploaded trough the submenu as we can see on the upload dialog of the new annotations on the Fig .5.

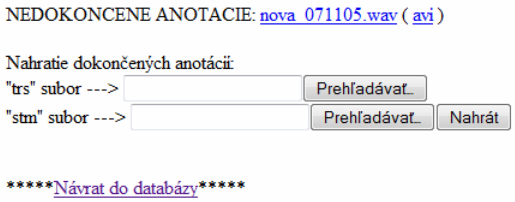

Fig. 5. The upload dialog for sending new annotations or to download the multimedia content of the annotations, which are in progress.

Finally the filenames of the annotations chosen for the upload are checked before the operation of the replacement is executed, and if the filenames are not identical with the expected ones, they are rejected. They are also more reasons for rejection of the uploaded files (they are blank, or no file was chosen, etc.) as we can see on the Fig 6.

The successful uploading leads to correcting the database, backuping old annotations, logging of the annotators actions and sending a new files list to the administrator, who could include them in the training or testing material of the next development of the automatic speech recognition engine sources or adapting the current language or acoustic models.

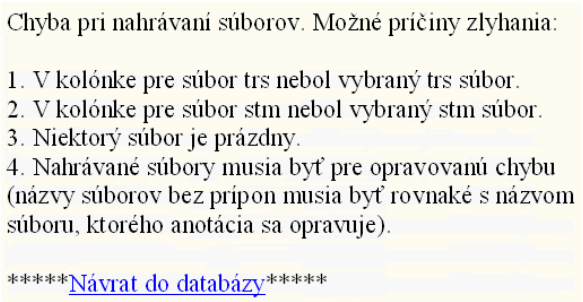

Fig. 6. Error screen presenting possible reasons, when the upload of the annotations will be rejected.

IV. Conclusion

The multimedia database interface presented in this paper is a very important management tool for annotation of the new multimedia content and correcting the presented one. Without this tool the work is very difficult for the database administrator, who is responsible on managing the work of all annotators. Especially if the annotator team is bigger or the corrections of the material is very time consuming.

The next big advantage of this approach is that every user (and also administrator) could access the whole database and the information about a possible new or unfinished work trough the web interface and then the worker enjoy flexibility in working location and hours, which means this is a possibility for teleworking or working at home (WAH) [11].

Acknowledgment

Research described in the paper was financially supported by the Ministry of Education of Slovak Republic under research projects AV 4/2016/08, the Slovak Grant Agency projects APVV-0369-07, APVV MobilTel No. APVT-20-029004, the EU FP7 project INDECT - No. 218086 and also by VEGA project No.1/4054/07.

References

[1] J. Juhár, S. Ondáš, A. Čižmár, M. Rusko, G.Rozinaj, R. Jarina, “Development of Slovak GALAXY/VoiceXML Based Spoken Language Dialog System to Retrieve Information from the Internet”, Proc. of ICSLP, Pittsburgh, USA, September 17-21, 2006, pp. 485-488.

[2] M. Rusko, S. Daržagín, M. Trnka, “SpeechDat-E telephone speech database as an important source for basic acoustic-phonetic research in Slovak”, in Proceedings of ICA 2004, Kyoto, Japan, 2004, pp. II-1676-II-1682.

[3] M. Pleva, J. Juhár, A. Čižmár, “Slovak broadcast news speech corpus for automatic speech recognition”. In: RTT 2007: Research in Telecomunication Technology 2007: 8th International Conference: Žilina - Liptovský Ján September 10-12, 2007. Žilina: ŽU, 2007. 4 s. ISBN 978-80-8070-735-4.

[4] S. Renals, D. Abberley, D. Kirby and T. Robinson, “Indexing and retrieval of broadcast news”, Accessing Information in Spoken Audio Speech Communication, Volume 32, September 2000, Pages 5-20.

[5] MSDN library, “Understanding SAMI 1.0”, Microsoft Corporation Originally published October 2001; updated February 2003, Available: http://msdn.microsoft.com/en-us/library/ms971327.aspx

[6] http://trans.sourceforge.net/en/presentation.php

[7] http://www.webaim.org/techniques/captions/windows/40_009a03.wmv

[8] http://webfilebrowser.sourceforge.net

[9] http://dev.mysql.com/downloads/

[10] http://www.php.net/downloads.php#v5

[11] http://en.wikipedia.org/wiki/Telecommuting