D9.6 ESTABLISHMENT OF THE INFRASTRUCTURE TO AUTOMATICALLY ANALYSE OTHER DATASETS

Revision: v1.0

Work Package	WP9
Task	T9.4
Due date	30/06/2022
Submission date	30/06/2022
Deliverable lead	Universität Hamburg
Version	1.0
DOI (latest version)	10.25592/uhhfdm.10264
Authors	Thomas Hanke (Universität Hamburg – UHH)
Reviewers	Mélanie Henault-Tessier (INT)

Abstract	This deliverable documents the installation of a tool chain for processing sign language data external to the project, mostly meant to be run on a high performance cluster.
Keywords	OpenPose, Media Pipe, High Performance Cluster, video processing, annotation, corpus

Grant Agreement No.: 101016982
Call: H2020-ICT-2020-2
Topic: ICT-57-2020
Type of action: RIA

Document Revision History

Version	Date	Description of change	List of contributors
V0.1	28/06/2022	First draft	Thomas Hanke (UHH)
V0.3	29/06/2022	Internal Reviews	Mélanie Henault-Tessier (INT)
V1.0	30/06/2022	Camera-ready submission DOI: 10.25592/uhhfdm.10265	Thomas Hanke (UHH)

DISCLAIMER

COPYRIGHT NOTICE

Project co-funded by the European Commission in the H2020 Programme		
Nature of the deliverable	**OTHER**	
Dissemination Level		
PU	Public, fully open, e. g., web	✓
CL	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to EASIER project and Commission Services	

* R: Document, report (excluding the periodic and final reports)
DEM: Demonstrator, pilot, prototype, plan designs
DEC: Websites, patents filing, press & media actions, videos, etc.
OTHER: Software, technical diagram, etc

Funded by the Horizon 2020 Framework Programme of the European Union

EXECUTIVE SUMMARY

This deliverable documents the installation of a tool chain for processing sign language data external to the project, mostly meant to be run on a high performance cluster.

Funded by the Horizon 2020
Framework Programme of the European Union

CONTENTS

Funded by the Horizon 2020
Framework Programme of the European Union

1 INTRODUCTION

It is one of the goals of EASIER to make it easier for projects not yet involving sign language technologies to profit from them. For this purpose, the project offers to process video data from outside the project on high performance clusters to get OpenPose[1] and Media Pipe[2] data from the videos in order to base time series analyses on these data. At the same time, data curation algorithms as developed in D6.6 are applied to any annotation data if supplied. In exchange, EASIER has a clearer view on what resources are available on European Sign Language to adapt tools where necessary.

At this point of time in the lifespan of the project it does not seem necessary nor advantageous to fully automate the whole process. Instead, only access to the data, the computationally heavy processing of the video data as well as the return of the results have been automated.

For processing the video data, we build on and have streamlined the use of a high-performance cluster at UHH. Tools for feeding into the process, collecting results as well as loadbalancing developed in the DGS-Korpus project (Hanke, 2019) have been generalized to the range of data to be expected.

Transferring source data between external projects and the EASIER processing site may mean huge file transfers the data owners may not be able to initiate. We therefore rely on a service called GigaMove[3] that RWTH Aachen makes available to all German academic institutions. The service allows the initiator to request data from someone else or to make data available to someone else with a limit of 100GB per file and up to 1000GB stored on the server for the initiator.

For starting the process, we foresee and exchange via email or videoconferencing to decide what tools can be applied to their specific data and to discuss further needs. Filling the web form for starting the process is only considered a formal confirmation of the discussion results. As has been outlined in D6.6, processing of data depends on the annotation conventions used. Again, manual intervention at the side of EASIER might be necessary to adapt the import process.

[1] https://github.com/CMU-Perceptual-Computing-Lab/openpose
[2] https://mediapipe.dev
[3] https://gigamove.rwth-aachen.de/

Funded by the Horizon 2020 Framework Programme of the European Union

2 ACCESS TO THE DATA

There are several projects out there providing services to linguistic data owners, mostly by providing a web interface to upload the data to be analysed (e.g. QUEST, Arkhangelskiy et al. (2021)). With sign language videos, the situation is not so straightforward due to the sheer size of video data in use in sign language corpus projects. Furthermore, videos showing subjects are automatically data with privacy constraints. Therefore, the process established differs substantially from those seen for text corpus analysis.

In the web form to initiate the process, the contributor declares to be the data holder, i.e. to have the right to make the data available to EASIER for processing, identifies a contact person with an email address in case of any questions, optionally a json or txt file specifying the structure of the data to be processed.

The txt file as a simple alternative to the json file contains one line per file in the dataset. That line is either a URL from where to download that file or just a file name.

If the text file contains a least one file name instead of a URL, the processor will initiate a GigaMove invitation to upload a zip archive of these files.

The data to be processed then consists of the files downloaded from the URLs provided as well as the files uploaded via GigaMove.

If the json file contains at least one video specification without a URL, the processor will initiate one or more GigaMove invitations to upload the missing files in zip archives. The size attributes determine how many invitations will be generated.

The uploads should then be done by sessionID so that the processing can start as soon as some sessions are completely available to the processor.

If no json or txt file is provided, the processor will initiate a GigaMove invitation to upload a zip archive of the complete dataset which basically limits this option to small datasets.

In the case that no json file was provided, the structure will be guessed from the input data, i.e. from file naming under the assumption that all files have the sessionID as its main component. Note that this will also assume that all recordings have the same camera perspective. If there are multiple video files for one recordings, they are assumed to be named in alphabetical order without gaps between the files.

The contributor should make the highest quality option available for the videos even if they are not suitable for any other processing due to their size.

However, if timelines of videos need to be manipulated (e.g. to keep synchrony between perspectives), then the upload should contain the best quality available with these corrections applied.

Funded by the Horizon 2020
Framework Programme of the European Union

3 JSON FILE FORMAT

3.1 JSON FILE SUBMITTED

The json file needs to have the following structure:

{ sessionID: { "videos": [...], "annotation": url } }

with a sessionID being any unique identifier and each of the video array elements being an object consisting of the following fields:

perspective: any text,

start: timecode,

end: timecode,

time base: optional denominator for the fractions of a second specified in timecode strings,

size: optional integer for file size in bytes,

url: where to download this file from, if missing, processor will generate a GigaMove upload invitation, for this, the size attribute is important,

file name: optional string used when no url is provided to generate an upload invitation,

participants: array of participantIDs,

timecodes can be either a text string such as "10:04:57:24" (hours:minutes:seconds:frames) or "10:04:57.480" (hours:minutes:seconds with the seconds be a float value) or an array [36297,24,50] as defined in D6.6: seconds, fractions numerator, fractions denominator. In the case of all-colons strings, the frames are just the numerator, with the optional time base value as the denominator. When no time base is specified, it is assumed to be identical to the frames per second value of the video file itself unless there are larger frames values specified. In that case, the denominator is assumed to be d = 2 ** ceiling(ld(max(frames)/fps)) with ld being the logarithm to the basis 2. (E. g. if the maximum frames count is 48 for a 25 fps video file, the time base is assumed to be 50.)

A perspective descriptor can just be any code string, it is of course optional when only one camera is used. In associating videos with the annotation, from any files sharing sessionID and perspective and having a timecode overlap, only the first video is taken in consideration.

The array of participantIDs provides codes for the participants being visible in the video file, using the same codes as in the annotation file.

In addition to the videos field, the session may have

annotation: file name of or url to an eaf or ilex or srt/vtt/ttml file

Funded by the Horizon 2020 Framework Programme of the European Union

3.2 DATA RETURNED

The data are returned to the submitter via the GigaMove service as zip archives. In addition to the newly created data, the archive contains a json file describing all contents processed and the results. If a json file was submitted, the file returned is that file with new information added as fields to the objects listed:

interchange: EASIER interchange format file,

openpose: filename,

mediapipe: filename

The interchange format file is only present if annotatation files have been submitted, the Open-Pose and Media Pipe files are only there if video data have been submitted.

More fields may become available later in the project.

Funded by the Horizon 2020
Framework Programme of the European Union

4 PROCESSING

4.1 PYTHON SCRIPTS

Python scripts have been written to process to create the json file from a txt file supplied, to retrieve data from URLs provided and to communicate with GigaMove to get access to the data. Other python scripts upload files to the HPC, start and monitor jobs there, and download the results. The final step uploads the data to be returned to GigaMove, generates a report and sends that to the data owner.

4.2 PROCESSING TIME

The time needed for processing data not only depends on its size, but also on the workload of the high performance cluster used. While it is theoretically possible that the processing time does not substantially exceed the duration of the video recordings submitted, a substantial overhead should be expected.

In the preparation of this deliverable, the process was simulated for the DGS part of the Dicta-Sign corpus (1080p50)[4] with an overall duration of slightly more than 48 hours, corresponding to roughly 9M frames or 900 hours of single-node processing time. With an average of 15 nodes available to the process at that particular time, it took about 2.5 days to complete. The major part of the processing time is occupied by OpenPose analysis. As OpenPose speed depends more on the number of persons visible on the frame than its spatial dimensions, increased processing times are to be expected for camera perspectives with a number of people visible.

4.3 DATA PRIVACY

With initiating the process, the data holder declares to have the rights to process the data in the way suggested by EASIER. EASIER ensures that the data is only processed by consortium members in countries completely complying with EU GDPR plus the provider of the GigaMove service which is a public institution in an EU member state as well. As EASIER processes the data as part of its research context, the processing does not fall under GDPR Data Processing. After processing the data and transmitting the results, EASIER will keep statistics on the data used in their research but not any person-related information, i.e. all video files will be deleted unless something else has explicitly been arranged. The statistical data collected will be deleted 10 years after the end of the project, i.e. in 2034.

[4] https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/

Funded by the Horizon 2020 Framework Programme of the European Union

REFERENCES

Arkhangelskiy, Timofey, Hanna Hedeland, and Aleksandr Riaposov (2021). “Evaluating and assuring research data quality for audiovisual annotated language data”. In: *Selected Papers from the CLARIN Annual Conference 2020: Virtual Event, 2020, 5-7 October*. Matsue, Japan, pp. 1–7. DOI: 10.3384/ecp1801.

Hanke, Thomas (2019). *Processing DGS-Korpus Data with OpenPose on the Hamburg High Performance Cluster*. Project Note AP04-2018-01. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Universität Hamburg. DOI: 10.25592/uhhfdm.835.

Funded by the Horizon 2020
Framework Programme of the European Union