

D5.2: Prepared platform and case studies

JSI-K4A, UPVLC

Distribution: Public

*trans*Lectures

Transcription and Translation of Video Lectures

ICT Project 287755 Deliverable D5.2



Project funded by the European Community under the Seventh Framework Programme for Research and Technological Development.



Project ref no.	ICT-287755
Project acronym trans Lectures	<i>transLectures</i>
Project full title	Transcription and Translation of Video Lectures
Instrument	STREP
Thematic Priority	ICT-2011.4.2 Language Technologies
Start date / duration	November 1, 2011 / 36 Months

Distribution	Public
Contractual date of delivery	October 31, 2012
Actual date of delivery	November 18, 2012
Date of last update	October 31, 2012
Deliverable number	D5.2
Deliverable title	Prepared platform and case studies
Type	Report
Status & version	Final version
Number of pages	15
Contributing WP(s)	WP5
WP / Task responsible	JSI-K4A, UPVLC
Other contributors	
Internal reviewer	Alfons Juan and Jorge Civera
EC project officer	Susan Fraser

The partners in *transLectures* are:

Universitat Politècnica de València (UPVLC)
XEROX Research Center Europe (XRCE)
Josef Stefan Institute (JSI)
Knowledge for All Foundation (K4A)
RWTH Aachen University (RWTH)
European Media Laboratory GmbH (EML)
Deluxe Digital Studios Limited (DDS)

For copies of reports, updates on project activities and other *transLectures* related information, contact:

The *transLectures* Project Coordinator
Alfons Juan, Universitat Politècnica de València
Camí de Vera s/n, 46018 València, Spain
ajuan@dsic.upv.es
Phone +34 699-307-095 - Fax +34 963-877-359

Copies of reports and other material can also be accessed via the project's homepage:
<http://www.translectures.eu>

© 2012, The Individual Authors

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

Executive Summary

This document contains details about WP5 task **5.2 Prepared platform and case studies** including all work that has been done within the scope of this task.

Table of Contents

- Executive Summary 3
- 1. Introduction 5
- 2. Architectural adaptations 5
 - 2.1 VideoLectures 5
 - 2.1.1 Changes to VideoLectures.NET code..... 5
 - Architectural changes to admin interface..... 5
 - New player 7
 - 2.1.2 Installation of server for storing transcriptions and translations 7
 - Enabling access to the server 8
 - 2.1.3 Creation of demo Web service for serving transcriptions 9
 - 2.2 Polimedia..... 9
 - 2.2.1 Production..... 9
 - 2.2.2 Processing 10
 - 2.2.3 Play 11
 - 2.3 Matterhorn 12
- 3. Operative plan..... 13
- 4. Conclusions 13
- 5. References..... 14
- A) Acronyms 15

1. Introduction

The objective of task 5.2 is to build and adapt a platform to support the services required for transLectures fulfilling the requirements described by the case studies of VideoLectures.NET, poliMedia and Matterhorn in WP 5.1. This is the intermediate task, which will be followed by the task 5.3 (Integration into the case studies) in which the models and tools developed in WP3 and WP4 will be incrementally integrated into VideoLectures.NET and poliMedia. Task 5.4 (Deployment and Matterhorn service development) will also be dependent on the results of this task.

This document described the services platform built in task 5.2 to support transLectures for VideoLectures.NET and poliMedia and also the architectural adaptations that has been made to support transLectures in VideoLectures.NET, poliMedia and Matterhorn.

Finally we will outline the next steps that we will take into the task 5.3.

2. Architectural adaptations

2.1 VideoLectures.NET

2.1.1 Changes to VideoLectures.NET code

Architectural changes to admin interface

Admin interface for VideoLectures.NET is currently being completely rewritten. With the new interface VideoLectures.NET editors/admins will be able to better control (view/edit/modify) settings, users, content's metadata and processes on the site. This will be especially useful for the support of transLectures features, which will be integrated in next step.

Let's consider the following admin screen for a lecture *cyberstat2012_parrondo_maxwell_demons*:

Filename	Status	Size	Type	Actions
cyberstat2012_parrondo_maxwell_demons_01.ss.jpg	✓	161.2 KB	Screenshot (.jpg, .gif, .png)	🗑️
cyberstat2012_parrondo_maxwell_demons_01.jpg	✓	6.2 KB	Thumbnail (.jpg, .gif, .png)	🗑️
cyberstat2012_parrondo_maxwell_demons_01.pdf	✓	3 MB	Slide Presentation (.pps, .ppt, .pptx, .pdf,...)	🗑️
cyberstat2012_parrondo_maxwell_demons_01.flv	✓	299.7 MB	Video (.flv, .wmv, .mp4, .webm, .mov, .rm, .avi, .qtv, .f4v, .ogv)	🗑️
cyberstat2012_parrondo_maxwell_demons_01.wmv	✓	226.4 MB	Video (.flv, .wmv, .mp4, .webm, .mov, .rm, .avi, .qtv, .f4v, .ogv)	🗑️
cyberstat2012_parrondo_maxwell_demons_01.mp4	✓	401.8 MB	Video - generic video source (not for distribution)	🗑️

Figure 1: Lecture admin screen for Media content

From here admin can see all the files/attachments available for a lecture, can select immediate actions like preview, can add new files or queue a new job for this lecture (for example video re-encoding, transcription, translation, ...).

One major change in VideoLectures.NET architecture is the introduction of process queues using the PostgreSQL database. When the admin selects the job, that needs to be executed asynchronously, this is done via the “semaphores” inside the VideoLectures.NET database. After this so-called “worker” processes start to execute jobs from the queue, returning results in predefined locations. These worker processes can execute on different machines across the network.

Using this principle a transcription and/or translation process will be initiated for a lecture. Worker process will send the video to a transcription Web service and collect the transcription when it’s done. The same will happen for a translation. These new actions have been added in the dropdown menu on button with a default action “Encode” (see picture above).

Because the process of transcription and/or translation could take some time to finish, the admin can monitor the status of submitted jobs via the Content/Lectures option in the admin interface.

Url_prefix	Type	Title	Date/time	Encoded	Thumbnail	Synced	Enabled
cyberstat2012_parrondo_maxwell_demons	vi	Maxwell demons, feedback control, and fluctuation theorems	2012-09-14	✘	✓	✓	✓
cyberstat2012_zecchina_message_passing	vi	A message-passing approach to stochastic optimization and inverse dynamical problems	2012-09-14	✘	✓	✓	✓
cyberstat2012_dall_asta_network_dynamics	vi	Activation and control of extreme trajectories in network dynamics	2012-09-14	✘	✓	✓	✓
cyberstat2012_todorov_optimal_control	vi	From linearly-solvable optimal control to trajectory optimization, and (hopefully) back	2012-09-13	✘	✓	✓	✓
cyberstat2012_braun_ortega_thermodynamics	vi	Thermodynamics as a theory of bounded rational decision-making	2012-09-13	✘	✓	✓	✓
pms2012_kotnjek_kiti	vi	Kiti v severemem Jadranu	2012-09-13	✘	✓	✘	✓
cyberstat2012_chertkov_fluid_mechanics	vi	Fluid Mechanics = "Single"-Particle Control: From Control to Turbulence and Back	2012-09-13	✘	✓	✓	✓
cyberstat2012_chernyak_stochastic_control	vi	Stochastic Control as a Non-Equilibrium Statistical Physics: Gauge Invariant Bellman Equation	2012-09-13	✘	✓	✓	✓
solomon_shmueli_count_data	vi	A Flexible Model for Count Data: The COM-Poisson Distribution	2012-09-13	✘	✓	✓	✓
cyberstat2012_tishby_bellman_equation	vi	How to discount Information: Information flow in sensing-acting systems and the emergence of hierarchies	2012-09-13	✘	✓	✘	✓
cyberstat2012_friston_free_energy	vi	Free energy and active inference	2012-09-13	✘	✓	✓	✓
cyberstat2012_delvenne_control_theory	vi	The thermodynamics of port-Hamiltonian systems	2012-09-13	✘	✓	✓	✓
cyberstat2012_brockett_attention_control	vi	Attention as a Design Criterion and Analysis Tool in Control Problems	2012-09-12	✘	✓	✓	✓
cyberstat2012_guerra_time_reversal_invariance	vi	Stochastic variational principles for dissipative and conservative systems	2012-09-12	✘	✓	✓	✓
cyberstat2012_hurtado_spontaneous_symmetry	vi	Spontaneous Symmetry Breaking at the Fluctuating Level	2012-09-12	✘	✓	✓	✓

Figure 2: Lecture admin screen for monitoring status

When the transcription/translation is done, the returned files are stored in Git repository under the **transgit** account. This repository is managed by partner XRCE on VideoLectures.NET’s new server for storing transcriptions and translations (see next chapter).

New player

To be compatible with new platforms like tablets, a new player was developed for VideoLectures.NET. It comes in two varieties:

- version supporting Flash
- version supporting HTML 5



Figure 3: Example of a new player appearance

Each version will support the required CC button for enabling subtitles. CC button in the lower right corner will show all available subtitles in original and translated languages in a pop-up menu. The user will select the language and the subtitles will be shown in the lower part of the player (overlaid on top of the video).

Player will check the availability of transcriptions and translations via Web service running on new VideoLectures.NET's server as described in deliverable D5.1^[2].

2.1.2 Installation of server for storing transcriptions and translations

For the purpose of storing and updating transcriptions and translations all partners on the transLectures project need to have access to a repository, where these transcriptions and translations will be stored. Each site (VideoLectures.NET and poliMedia separately) will have one such repository available on a local server.

For VideoLectures.NET a new computer server was bought in June 2012 for this purpose. It is a Supermicro system with the following components:

- Supermicro SuperServer 6017R-WRF in rack mount (1U) chassis
- 2x Intel Xeon E5-2630 2.3Ghz processors (2x 6 cores/12 threads)
- 32GB DDR3-1333 ECC RAM (8x 4GB)
- System disk: Seagate 300GB 15000rpm SAS2
- Storage disks: 3x Hitachi 3TB 7200rpm in RAID-5 (6 TB)

- Adaptec RAID 6405 Controller
- DVD-RW drive: NEC AD-7710H

On this server the following software was installed:

- **Ubuntu 12.04 amd64 server** with specially built kernel including the patch for Unionfs (see <http://unionfs.filesystems.org/>)
The inclusion of Unionfs patch in kernel enables merging the contents of several directories in one mounted directory, while keeping their physical content separate. With the kernel modification the need for context switching between kernel and user-space is greatly reduced, so disk access via unionfs produces significantly less overhead.
- Several **tools for speech & text processing**
 - o AK Toolkit – A toolkit for building and using Hidden Markov Models (HMMs). Originally developed for handwritten text recognition using Bernoulli HMMs, it also implements conventional mixtures of diagonal Gaussian distributions, and can be used for other tasks as Automatic Speech Recognition (ASR). AK consists of a main library, which implements the core features of the toolkit, the tools which allow the interaction with the library using a command line interface, and several scripts (Python, AWK) and tools, which are independent or only partially dependent of the library, used for preparing data and manipulating models. Important features in AK are:
 - Maximum likelihood parameter estimation of embedded HMMs using the Baum-Welch or Viterbi algorithms.
 - Continuous recognition using conventional language models (finite state models or ARPA n-gram models) and the most common pruning techniques (beam search, histogram pruning and look-ahead).
 - Adaptation of Gaussian HMMs using MLLR.
 - Support for binary and real feature sequence vectors.
 - Adaptation of real feature sequences using Constrained MLLR and fast Vocal Tract Normalization.
 - Support for tied parameters, triphon HMMs and cross word modeling.
 - o Apache Tika - for detecting and extracting metadata and structured text contents from various documents
 - o Tesseract - a commercial quality OCR engine for extracting text from images
 - o Html2text, pdftotext, ...
 - o Converters for various subtitle formats (srt, trs, dfxp, ...)

Enabling access to the server

Partners on the project will have remote access to the repository via Git over SSH. For this purpose an account with the name **transgit** was created. Another account was created for partner XRCE, who will install and prepare the necessary environment and Git repository under the **transgit** account.

Partners will then provide their public SSH keys, which will be stored in the **transgit** account thus enabling them access to the Git repository via remote git command.

Josef Stefan Institute's firewall was also modified to enable remote access to the transLectures server on required ports.

2.1.3 Creation of demo Web service for serving transcriptions

For preparing and testing the needed modifications for VideoLectures.NET we created the following:

- Local copy of current transcriptions repository
- Demo Web service implementing the following interfaces:
 - o GET /status
 - o GET /dfxp

Using this demo Web service (a subset of Web service described in D5.1^[2]) we were able to use the local copy of current transcriptions repository as read-only repository for testing purposes (for example serving available transcriptions to the new player).

2.2 poliMedia

Under the terms of the project, we have modified the workflow of poliMedia Service for supporting the architectural constraints derived from WP 5.1. These changes are described in this deliverable.

The architecture of poliMedia Service is a three step model, from the recording to the user consumption. These three phases are:

- **Production**, in which a teacher is recorded and we get the raw sources of audio, video and metadata
- **Processing**, in which the raw sources are processed and we obtain the final video. This video then is distributed to the streaming server
- **Play**, in which users access the content and metadata in a usable way

Following the case studies of deliverable D5.1^[2], many developments and changes are required to implement the use of transLectures with poliMedia. We will describe these changes in the next sections.

2.2.1 Production

- Teachers are encouraged to add slides and additional references and give it to the technicians in the recording session.
- We have modified the admin interface for poliMedia to get track of this additional content, as seen in the picture below.
- Audio, Video, Metadata and additional content is packed into a Matterhorn media package^[1] to allow easy ingesting to the transLectures Web Services.

polimedia
Admin
Cerrar Sesión

Nueva de prueba
Nueva grabación
Envío OA's pendientes
Listado polimedias
Catálogo

Editar bloque "Búsqueda de Bibliografía en ISBN"

[Bloques](#) > Editar

Operador:

Ruta:(Obligatorio)

M03/
Orden:

Título: (Obligatorio)

Materia:

Metadatos:

Profesor

E-mail: Nombre:

Apellidos:

Teléfono:

Cargo:

Entidad:

Es un objeto de aprendizaje
(Enviado a Riunet: 31/12/2009)
(Estado riunet: el usuario anonimo tiene permiso en el OA)

Estadísticas visibles

Fecha:

Permisos

Acceso:

Asignatura PoliformaT:

Ej: ALU_XXXX_2011

Visibilidad

Visible No se muestra en el catálogo

Contenidos

Módulo:

Curso:

Versión:

Figure 3: Polimedia admin interface

2.2.2 Processing

- We have installed a new server to implement the transLectures services in the poliMedia environment. This server is called *fuster.upv.es* and holds a copy of the current transcriptions and translations, and also the current transLectures Video Player

We have also implemented several web services, as planned in WP5.1. A detailed explanation of these services is available on that deliverable D.51, and. Here is a short list on those services:

- **ingest** , a POST service that reads an id and a Matterhorn media package and calls the transLectures tools to get the transcriptions and translations
- **status**, a GET service to know the available transcriptions and translations for an id
- **dfxp**, a GET service to get the DFXP file for an id and a language tag
- **update**, a POST service for the transLectures video player to update the current transcription/translation

2.2.3 Play

- We have modified the Polimedia video player, as shown in the mockups of D5.1^[2] for being able to check the available transcriptions and translations from the status interface of server *fuster.upv.es* and showing up a list of available transcriptions and translations
- We have modified the poliMedia video player for being able to get the DFXP translations from server *fuster.upv.es* and displaying it.

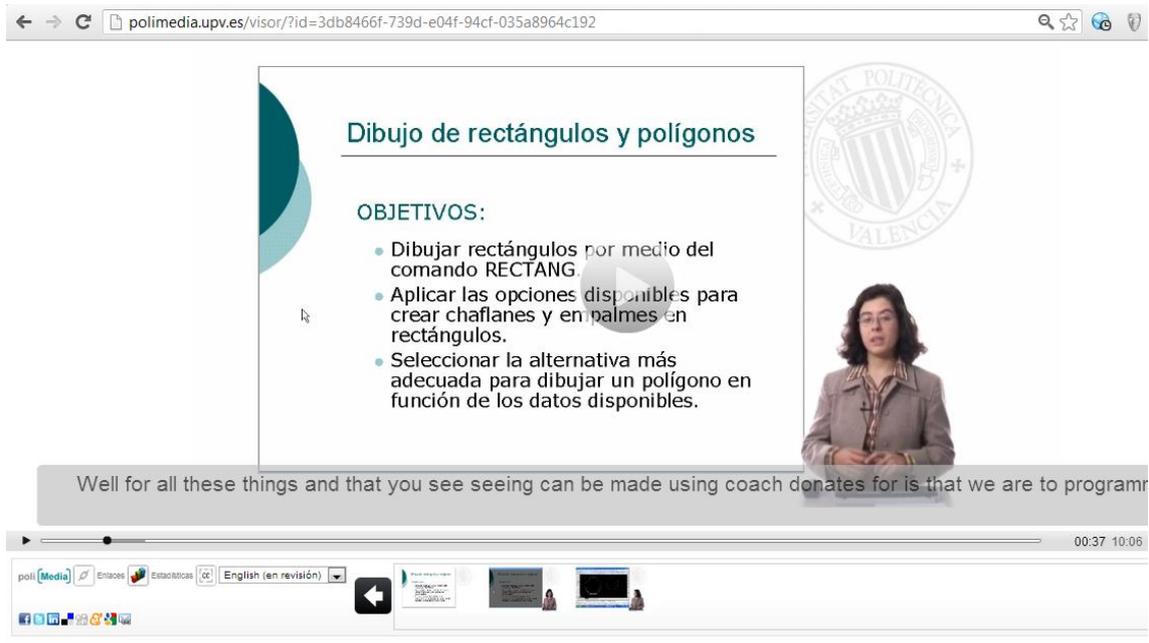


Figure 4: poliMedia player with the new subtitles and edit options

- We have added an “Edit” button to the poliMedia Player to jump to the transLectures video player to edit a transcription/translation. This work is closely related with the developments done in WP4, and is the gateway through the use on the intelligent interaction phase by the users.



Figure 5: Translectures player for editing & intelligent interaction

2.3 Matterhorn

Matterhorn is currently providing a template for a rudimentary speech recognition API, where an outside (commercial) Web service is contacted. Service accepts video (or audio only), processes it and returns the result of transcription. As this part of the Matterhorn code is not open source, the base API functions and workflows needs to be re-implemented.

To implement this functionality in Matterhorn, we have worked on the following parts of the code:

- Matterhorn speech recognition API:
 - o Implement functions **transcribe**, **isLanguageAvailable** and **getSupportedLanguages**
 - o Define function **translate**
- Matterhorn workflows:
 - o Workflow which includes transcription via transLectures Web service (Transcription Workflow Operation Handler)
- Matterhorn admin GUI:
 - o Hold option for transcription review

While this work should be finished at this time, as planned on the project, Matterhorn development has slowed in the last year, and we are expecting for the release of version 1.4 of Matterhorn, that was planned for July 2012 and isn't already finished. This release provides

the Episode Service on which we have to rely to implement transLectures on Matterhorn, and so this part should be considered already as a work in progress.

3. Operative plan

Work in the next period will be focused on incremental integration of models and tools developed in WP3 and WP4 into VideoLectures.NET and poliMedia, as planned in the DoW. We will specially take care consider the results from WP3, WP4 and WP6 to refactor the developments made. According to them, some optimisations will be done to maximise scalability and response effectiveness will also be implemented.

Work on Matterhorn will continue after the release of version 1.4, which is now planned for December 2102 by creating the following services:

- Implementing all functions for speech recognition API
- Creating workflows for transcriptions/translations via transLectures Web service
- Adding hold options in Matterhorn admin GUI for review/edit transcriptions and translations

4. Conclusions

This document presents both the platform we have deployed to do the integration of the transLectures software into the case studies and also the architectural adaptations needed to support that software according to the requirements defined on those case studies.

The main result of work package 5.2 is that now we have a service platform, compatible with VideoLectures.NET and poliMedia in which we can deploy the transLectures software and in which we can test and evaluate the transcriptions and translations provided by the project.

5. References

1. "Annex I - Description of Work" of Grant agreement for project transLectures, version 2011-07-11
2. Deliverable D5.1: Case study scenario definition [transLectures-D5.1-31May2012.pdf](#)
3. transLectures project Web page: <http://translectures.eu>
4. VideoLectures.NET Web page: <http://videlectures.net>
5. Internal poliMedia Web page at UPV
6. Matterhorn Web page: <http://opencast.org/matterhorn/>

[All webpages accessed on 31. October 2012]

A) Acronyms

UPVLC	Universitat Politècnica de València
XRCE	XEROX Research Center Europe
JSI	Josef Stefan Institute
K4A	Knowledge for All Foundation
RWTH	RWTH Aachen University
EML	European Media Laboratory GmbH
TTML	Timed Text Markup Language format/standard
DFXP	Distribution Format EXchange Profile
WebVTT	Web Video Text Tracks format/standard
FLV, MP4, WMV	Different video formats
MP3, AAC, WMA	Different audio formats
CBR	Constant BitRate
VBR	Variable BitRate
H.264	MPEG-4 part 10 or AVC (Advanced Video Coding) standard for video compression
CC	Closed Captioning
API	Application Programming Interface
SOA	Service-Oriented Architecture
OSGI	Open Services Gateway Initiative framework for Java
WCMS	Web Content Management System
LMS	Learning Management System
OCR	Optical Character Recognition
SOAP	Simple Object Access Protocol for implementation of Web Services
REST	REpresentational State Transfer, a style of software architecture for distributed system
JSON	JavaScript Object Notation, lightweight text-based open standard
ATOM	XML language used for Web feeds
RSS	RDF Site Summary Web feed format