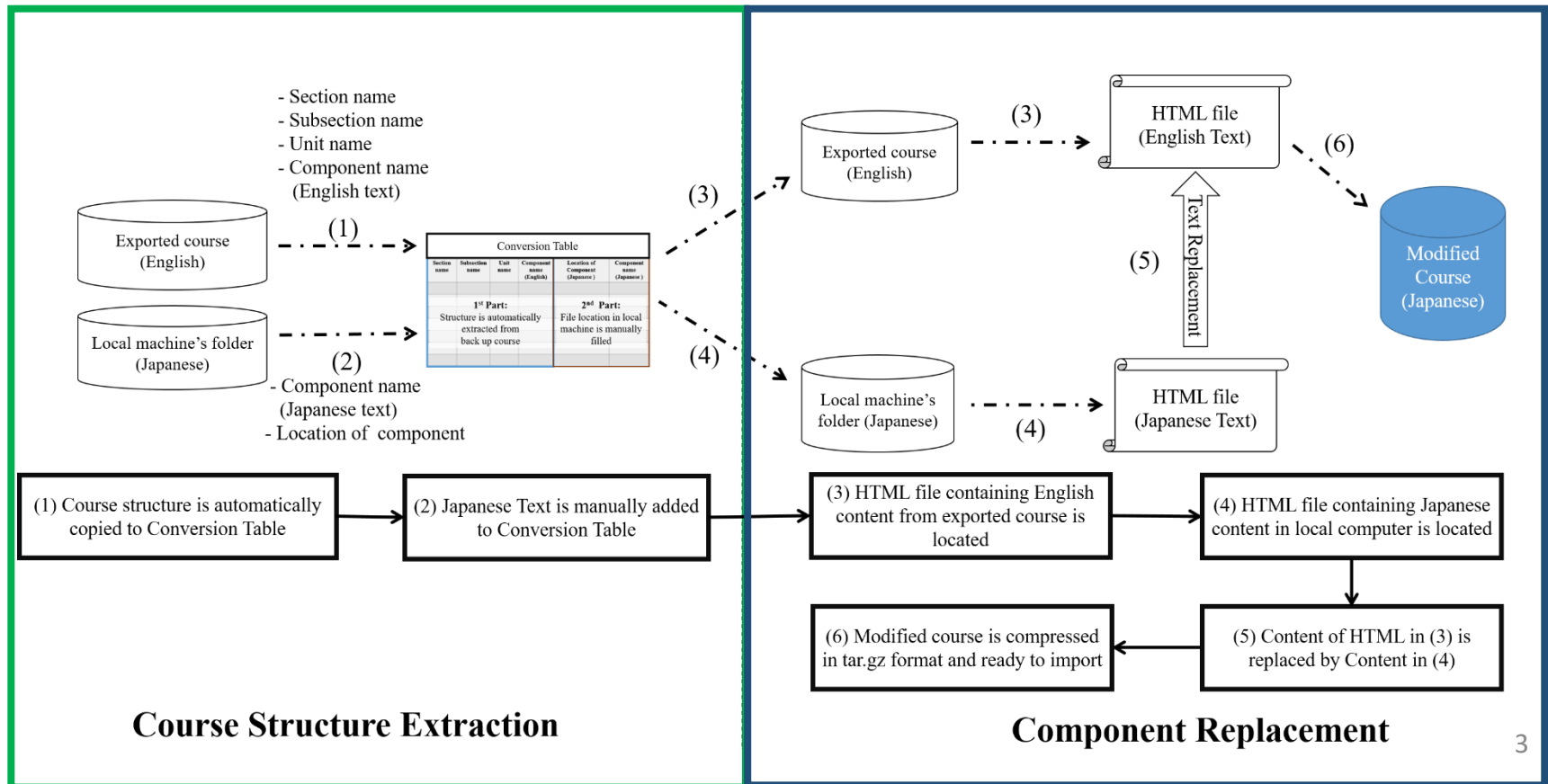# Text Modification Tool Tutorial

Contents

- How the tool work (slide 3-9)
- Tutorial (slide)
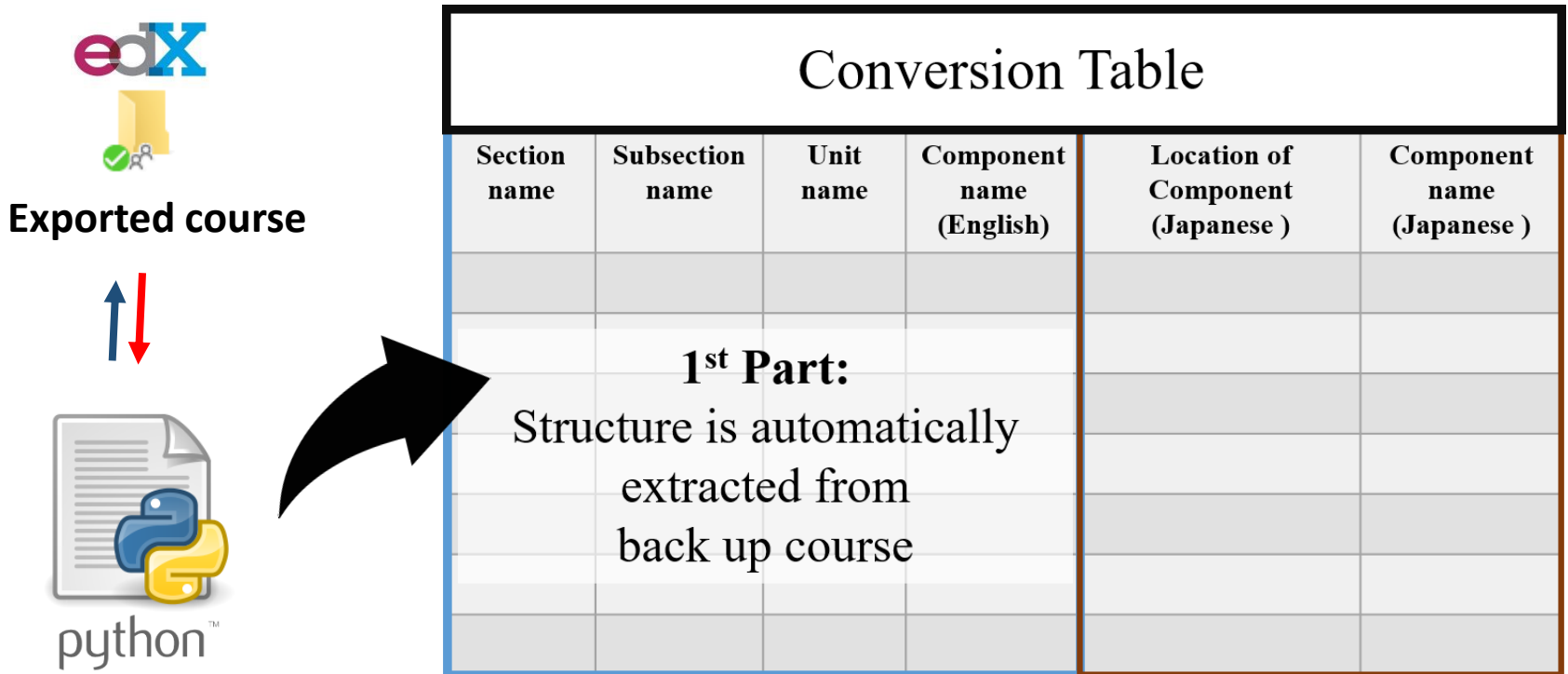
Overview of Text Modification Tool

- Consists of 2 subprocess

Course Structure Extraction

Component Replacement

3

Course Structure Extraction-1

- Course structures (**name**) are **copied** into the table

Exported course

Conversion Table					
Section name	**Subsection name**	**Unit name**	**Component name (English)**	**Location of Component (Japanese)**	**Component name (Japanese)**
1st Part: Structure is automatically extracted from back up course					

Course Structure Extraction-2

- Copy translated Files and location to the table

Conversion Table					
Section name	Subsection name	Unit name	Component name (English)	Location of Component (Japanese)	Component name (Japanese)
				2ⁿᵈ Part: File location in local machine is manually filled	

Translated text folder
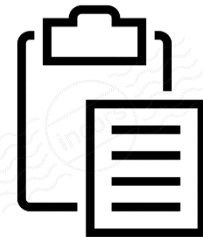
Copy and Paste

Example of Conversion Table

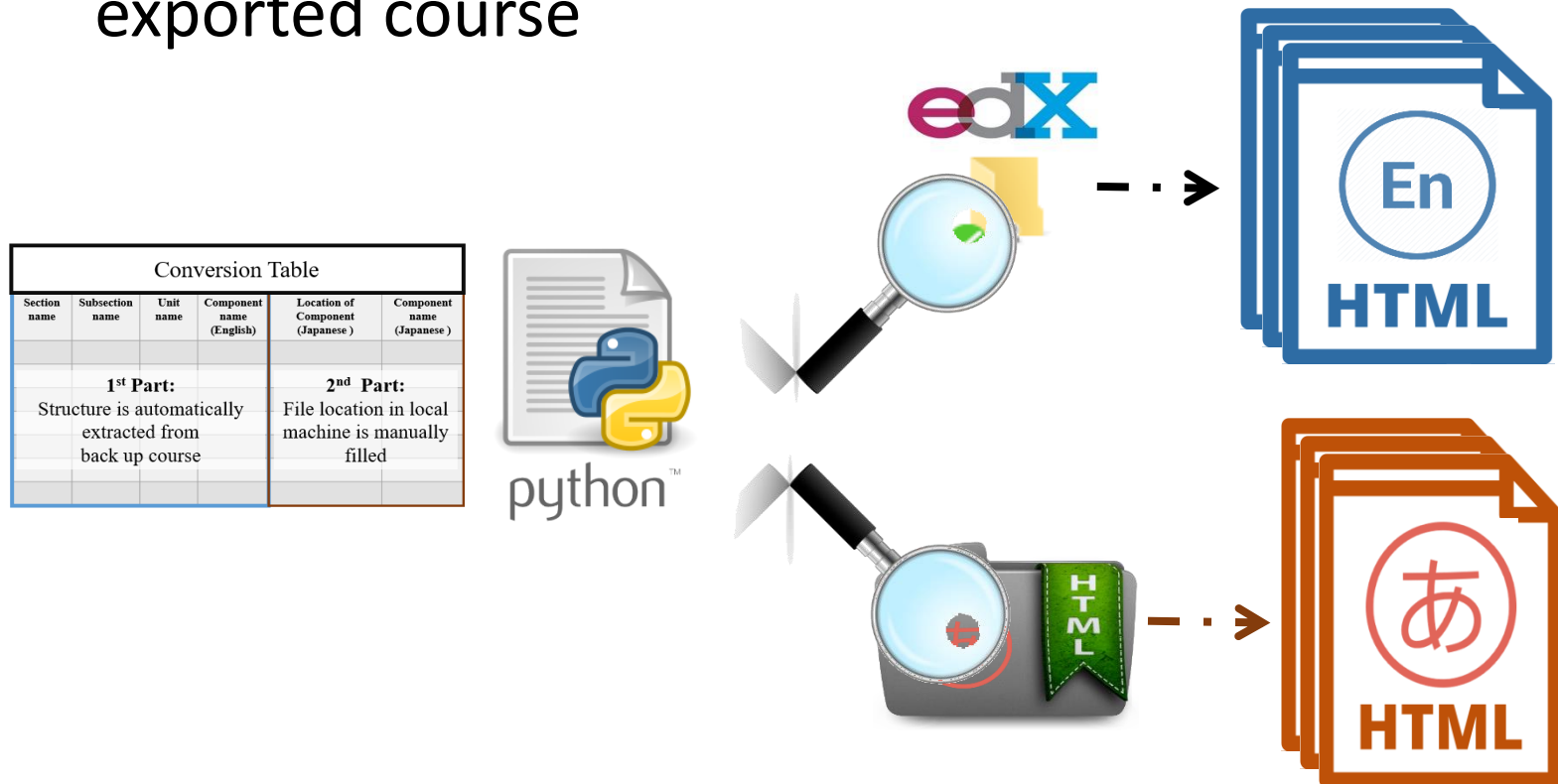section subsection unit Text component

Local directory HTML filename

no	section	subsection	unit	component_name	file_loc	file_name
143	Content Presentation	Overview	Standalone Slideshows	Resources	/5) content presentation/1) Overview/unit copy 6	comp4
144	Content Presentation	Overview	Standalone Slideshows	Pros and Cons	/5) content presentation/1) Overview/unit copy 6	comp5
145	Content Presentation	Overview	Videos: Screencasts	Videos: Screencasts	/5) content presentation/1) Overview/unit copy 7	comp1
146	Content Presentation	Overview	Videos: Screencasts	Description	/5) content presentation/1) Overview/unit copy 7	comp2
147	Content Presentation	Overview	Videos: Screencasts	Resources	/5) content presentation/1) Overview/unit copy 7	comp3
148	Content Presentation	Overview	Videos: Screencasts	Pros and Cons	/5) content presentation/1) Overview/unit copy 7	comp4
149	Content Presentation	Overview	Videos: Classroom Capture	Videos: Classroom Capture	/5) content presentation/1) Overview/unit copy 7	comp5
150	Content Presentation	Overview	Videos: Classroom Capture	Description	/5) content presentation/1) Overview/unit copy 8	comp1
151	Content Presentation	Overview	Videos: Classroom Capture	Resources	/5) content presentation/1) Overview/unit copy 8	comp2
152	Content Presentation	Overview	Videos: Classroom Capture	Pros and Cons	/5) content presentation/1) Overview/unit copy 8	comp3
153	Content Presentation	Overview	Videos: Studio or Green Screen	Videos: Studio or Green Screen	/5) content presentation/1) Overview/unit copy 8	comp4
154	Content Presentation	Overview	Videos: Studio or Green Screen	Description	/5) content presentation/1) Overview/unit copy 8	comp5
155	Content Presentation	Overview	Videos: Studio or Green Screen	Resources	/5) content presentation/1) Overview/unit copy 9	comp1
156	Content Presentation	Overview	Videos: Studio or Green Screen	Pros and Cons	/5) content presentation/1) Overview/unit copy 9	comp2
157	Content Presentation	Overview	Videos: Lightboard	Videos: Lightboard	/5) content presentation/1) Overview/unit copy 9	comp3
158	Content Presentation	Overview	Videos: Lightboard	Description	/5) content presentation/1) Overview/unit copy 9	comp4
159	Content Presentation	Overview	Videos: Lightboard	Resources	/5) content presentation/1) Overview/unit copy 9	comp5
160	Content Presentation	Overview	Videos: Lightboard	Pros and Cons	/5) content presentation/1) Overview/unit copy 10	comp1
161	Content Presentation	Overview	Videos: Location Shooting	Videos: Location Shooting	/5) content presentation/1) Overview/unit copy 10	comp2
162	Content Presentation	Overview	Videos: Location Shooting	Description	/5) content presentation/1) Overview/unit copy 10	comp3

1st Part: automated task   2nd Part: manual task
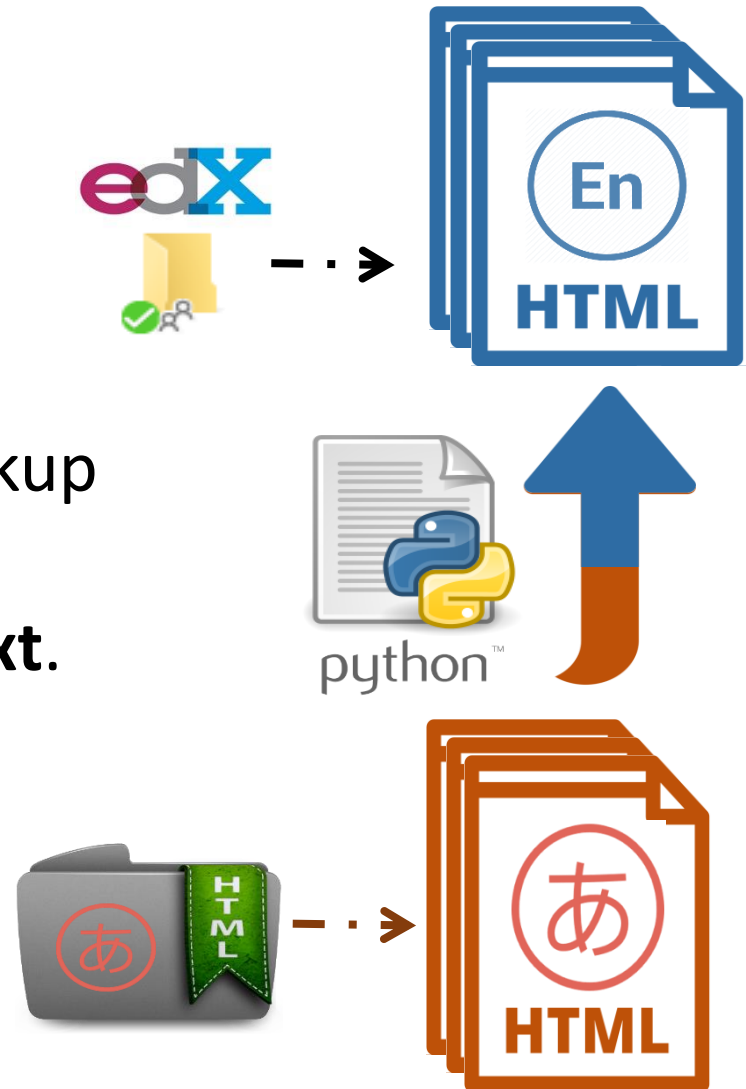
Component replacement-1

- Tool locates target HTML files (**English text**) in exported course

Conversion Table					
Section name	Subsection name	Unit name	Component name (English)	Location of Component (Japanese)	Component name (Japanese)
	1st Part: Structure is automatically extracted from back up course			2nd Part: File location in local machine is manually filled	

- Tool locates desired HTML files (**Translated text**) in local directory

Component replacement-2

- **English text blocks** in backup course will be iteratively replaced by **translated text**.

Component replacement-3

- Course with translated text is compressed in **tar.gz (edX's uploadable format)**

Output file

Use **edX's import function** to upload the modified course online
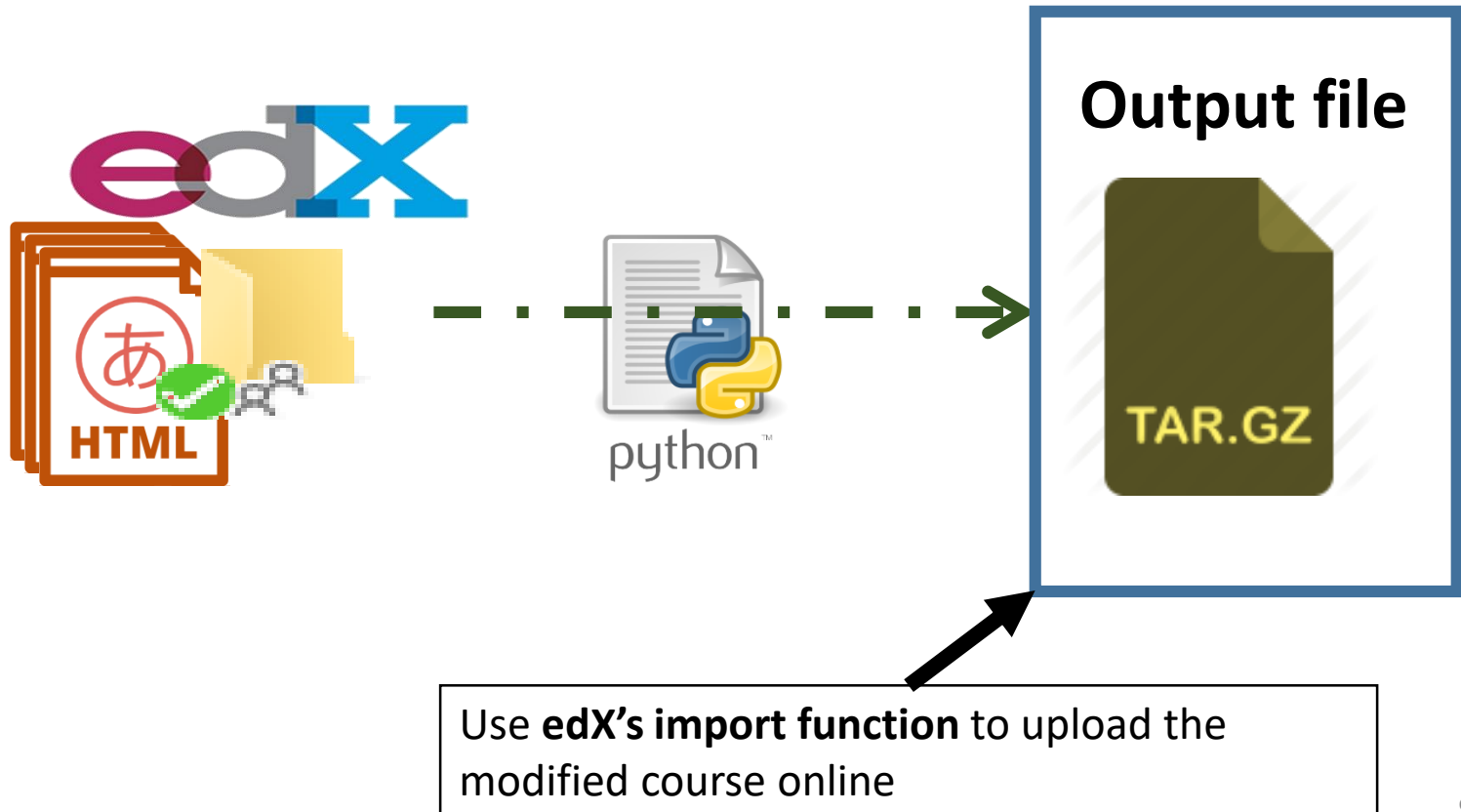
Workspace folder

- Python Dependence distribution
  - lxml, xlrd, xlwt, beautifulsoup4
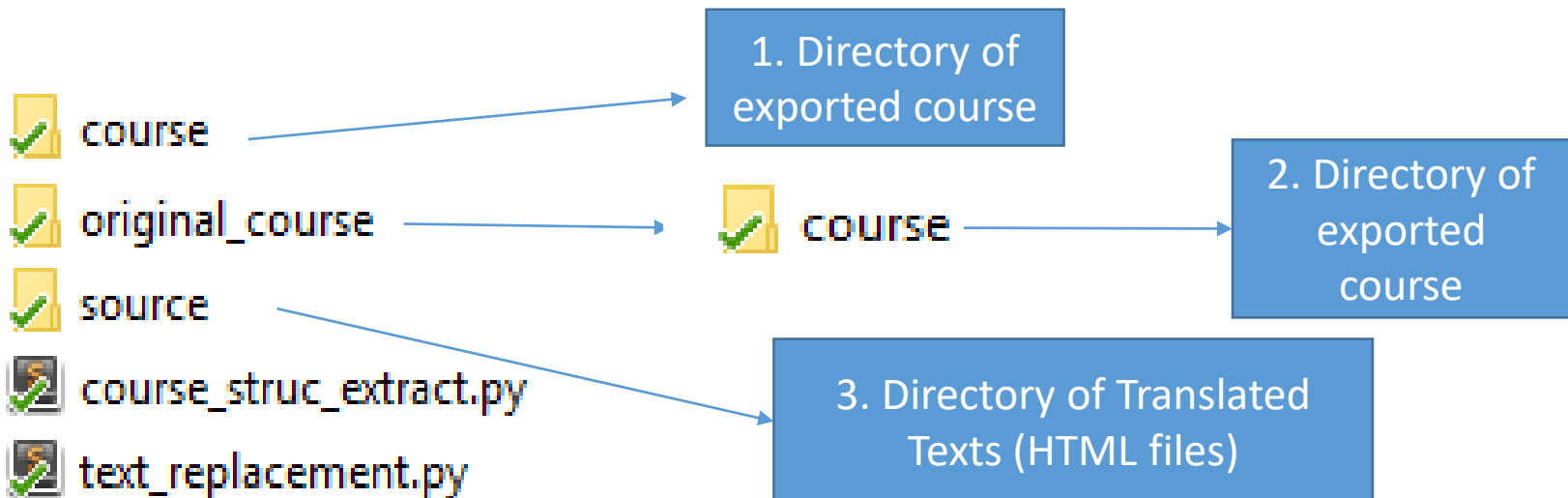- List of scripts

course_struc_extract.py

text_replacement.py

Course structure extraction: generates an excel file 'conversion table'

Component replacement: replaces text components to exported course

Files and Folders preparation

1. Unzip an exported edX course

    1.1 Rename folder of exported course as 'course'

2. Create 'original_course' folder

    2.1 Copy the 'course' folder into this directory

3. Create 'source' folder

    3.1 Move all files of translated text(HTML file) to 'source'

course

1. Directory of exported course

original_course → course → 2. Directory of exported course

source

course_struc_extract.py

text_replacement.py

3. Directory of Translated Texts (HTML files)

Example of exported course

Offline course structure

course →

original_course → course →

source

course_struc_extract.py

text_replacement.py

about
assets
chapter
course
drafts
html
info
policies

problem
sequential
static
tabs
vertical
video
word_cloud
course.xml
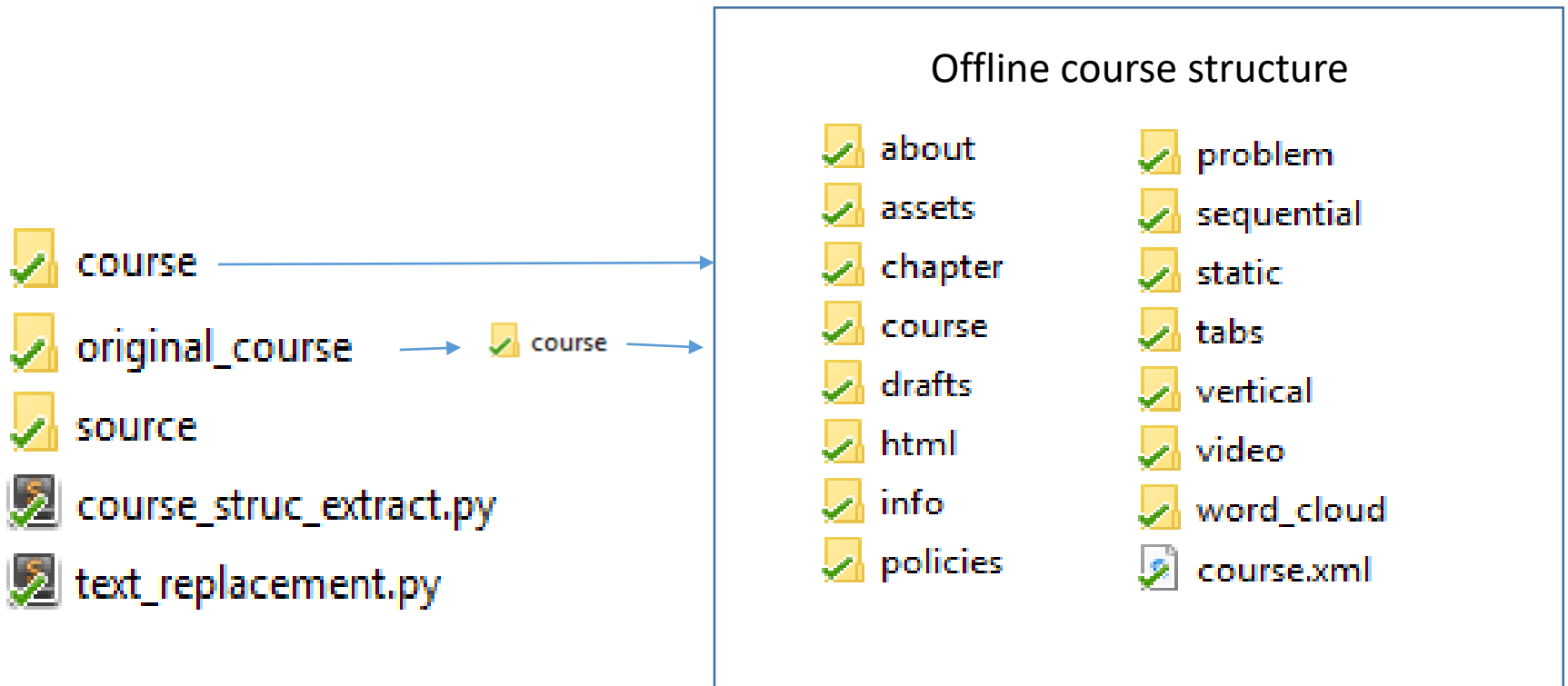
How to run the program

1. Run a python script   `python course_struc_extract.py`

Program show list of
extracted course contents

How to run the program

1. Run a python script 'course_struc_extract.py'
   - excel file 'conversion_table.xlm' will be created in the work space folder
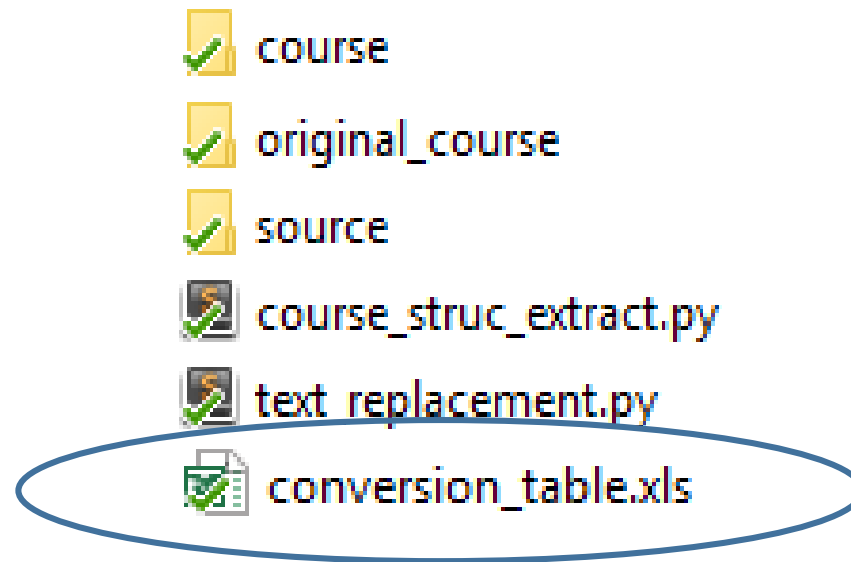
How to run the program

2. Open the conversion_table.xlm

- name of Section, Subsection, Unit and text components are automatically input in the table

no	section	subsection	unit	component_name
143	Content Presentation	Overview	Standalone Slideshows	Resources
144	Content Presentation	Overview	Standalone Slideshows	Pros and Cons
145	Content Presentation	Overview	Videos: Screencasts	Videos: Screencasts
146	Content Presentation	Overview	Videos: Screencasts	Description
147	Content Presentation	Overview	Videos: Screencasts	Resources
148	Content Presentation	Overview	Videos: Screencasts	Pros and Cons
149	Content Presentation	Overview	Videos: Classroom Capture	Videos: Classroom Capture
150	Content Presentation	Overview	Videos: Classroom Capture	Description
151	Content Presentation	Overview	Videos: Classroom Capture	Resources
152	Content Presentation	Overview	Videos: Classroom Capture	Pros and Cons
153	Content Presentation	Overview	Videos: Studio or Green Screen	Videos: Studio or Green Screen
154	Content Presentation	Overview	Videos: Studio or Green Screen	Description
155	Content Presentation	Overview	Videos: Studio or Green Screen	Resources
156	Content Presentation	Overview	Videos: Studio or Green Screen	Pros and Cons
157	Content Presentation	Overview	Videos: Lightboard	Videos: Lightboard

How to run the program

2. Open the conversion_table.xlm
- manually add the path location of translated texts (HTML files) contained inside 'source' folder
- For example: File of translated text 'Resources' components is located
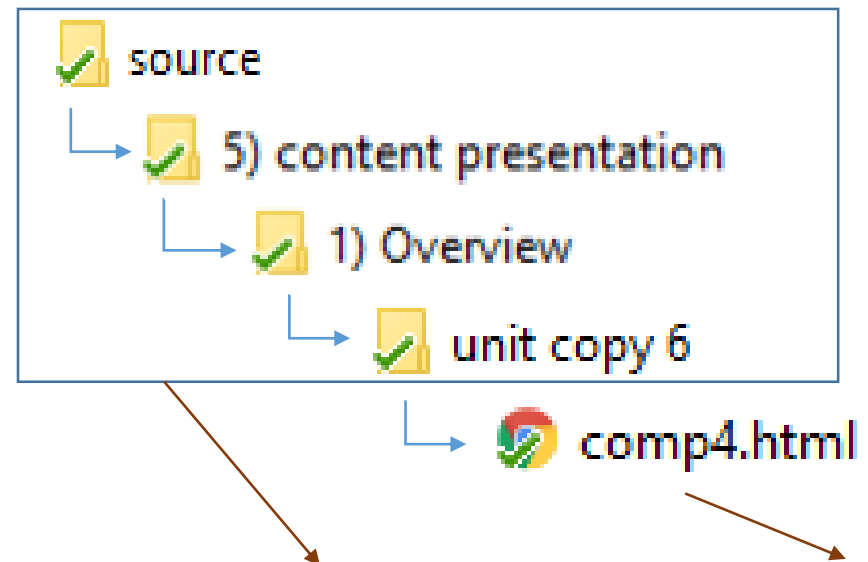
section	subsection	unit	component_name	file_loc	file_name
Content Presentation	Overview	Standalone Slideshows	Resources	/5) content presentation/1) Overview/unit copy 6	comp4

How to run the program

3. Run a python script
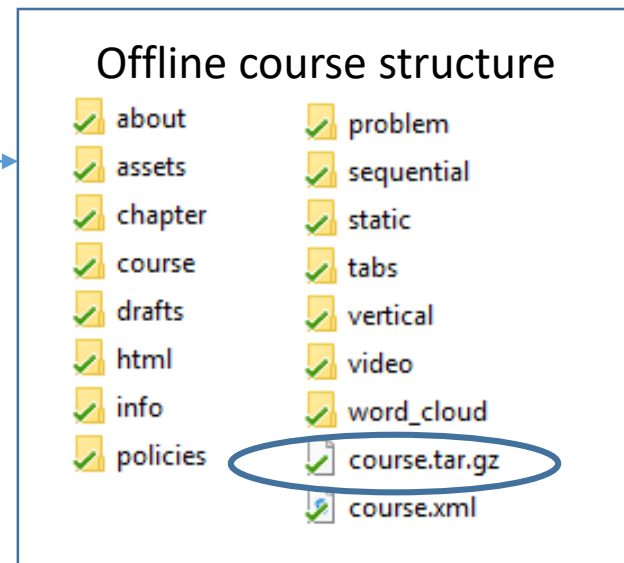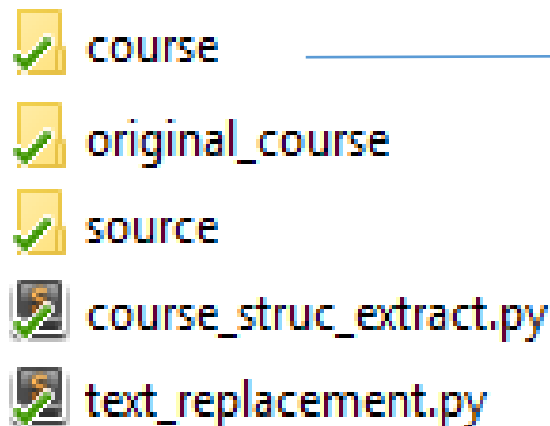
```
python text_replacement.py
```

List of replaced text

```
------------------- Eng-to-Jap text replacing is complete ----------------------
--------------------------------------------------------------------------------

----------------- begin replacing HTML component < Choosing Assessment Scoring Method > -----------------------
Found Sections: <Assessment> at filename --> 0977e2c4936b421ca310eea3a71dd8d7.xml
Found Subsections: <Implementation: Constructed-Response> at filename --> af0013825d7848e1943a0f87dd1e06fb
Found Units: <Constructed-Response Grader Options> at filename --> e33f2b25ebe1472288e1e6830e39b4f0
index number 123.0 Found HTML: <Choosing Assessment Scoring Method> at filename --> ec6df0298ebd4ed58271b53c91c1f7ba.xml
no-figure text content
------------------- Eng-to-Jap text replacing is complete ----------------------
--------------------------------------------------------------------------------

----------------- begin replacing HTML component < Constructed-Response Question Selection > -----------------------
Found Sections: <Assessment> at filename --> 0977e2c4936b421ca310eea3a71dd8d7.xml
Found Subsections: <Implementation: Constructed-Response> at filename --> af0013825d7848e1943a0f87dd1e06fb
Found Units: <Constructed-Response Question Selection> at filename --> 696562e4b1eb464a966e967d59a39538
index number 124.0 Found HTML: <Constructed-Response Question Selection> at filename --> 1a9ba95b6ec248de8688c1852df67ab0.xml
no-figure text content
------------------- Eng-to-Jap text replacing is complete ----------------------
--------------------------------------------------------------------------------

----------------- begin replacing HTML component < How to Identify Which Constructed-Response Question Type to Use > -----------------------
Found Sections: <Assessment> at filename --> 0977e2c4936b421ca310eea3a71dd8d7.xml
Found Subsections: <Implementation: Constructed-Response> at filename --> af0013825d7848e1943a0f87dd1e06fb
Found Units: <Constructed-Response Question Selection> at filename --> 696562e4b1eb464a966e967d59a39538
index number 125.0 Found HTML: <How to Identify Which Constructed-Response Question Type to Use> at filename --> fd9aba42b0104542b5a58cc0eb7f09db.xml
figure sources are all modified
------------------- Eng-to-Jap text replacing is complete ----------------------
--------------------------------------------------------------------------------

----------------- begin replacing HTML component < Challenges with Constructed-Response Questions > -----------------------
Found Sections: <Assessment> at filename --> 0977e2c4936b421ca310eea3a71dd8d7.xml
Found Subsections: <Implementation: Constructed-Response> at filename --> af0013825d7848e1943a0f87dd1e06fb
Found Units: <Constructed-Response Question Selection> at filename --> 696562e4b1eb464a966e967d59a39538
index number 126.0 Found HTML: <Challenges with Constructed-Response Questions> at filename --> 2cf05193e6ab4018a63bae9541b4aed2.xml
no-figure text content
------------------- Eng-to-Jap text replacing is complete ----------------------
--------------------------------------------------------------------------------

----------------- begin replacing HTML component < Self-Assessment-Implementation: Constructed-Response > -----------------------
Found Sections: <Assessment> at filename --> 0977e2c4936b421ca310eea3a71dd8d7.xml
Found Subsections: <Implementation: Constructed-Response> at filename --> af0013825d7848e1943a0f87dd1e06fb
Found Units: <Self-Assessment-Implementation: Constructed-Response> at filename --> 9db548e8f64f472d81f94e0100b9b88d
index number 127.0 Found HTML: <Self-Assessment-Implementation: Constructed-Response> at filename --> 5039ce9e420c4cd0b007de2ec938d743.xml
no-figure text content
------------------- Eng-to-Jap text replacing is complete ----------------------
--------------------------------------------------------------------------------
```

contents are not replaced. No translated text on the list

```
------------- no translated text file & directory :skipped HTML component  <  > ------------------
------------- no translated text file & directory :skipped HTML component  < Contents-Content Presentation Overview > -----------------
------------- no translated text file & directory :skipped HTML component  < Start with Small Bites > ------------------
------------- no translated text file & directory :skipped HTML component  < Benefits of Working with Small Bites > -----------------
------------- no translated text file & directory :skipped HTML component  < Reference > ------------------
```

How to run the program

4. If program is successfully executed, <u>ready-to-upload course (course.tar.gz)</u> is created inside the course folder

course

original_course

source

course_struc_extract.py

text_replacement.py

Offline course structure

about	problem
assets	sequential
chapter	static
course	tabs
drafts	vertical
html	video
info	word_cloud
policies	course.tar.gz
	course.xml

Notice

- This script is only tested with 1 edX course. Therefore, unexpected error should happen while using the program.

- There is a bug after uploading course back to edX. It is found that some text components disappeared in edX page. In our test, 1 component is missing.

- Some HTML component listed in 'error_files.txt' will not have figures showed properly in edX. This happens when original component (Eng) and translated component (local language) have different number of figures