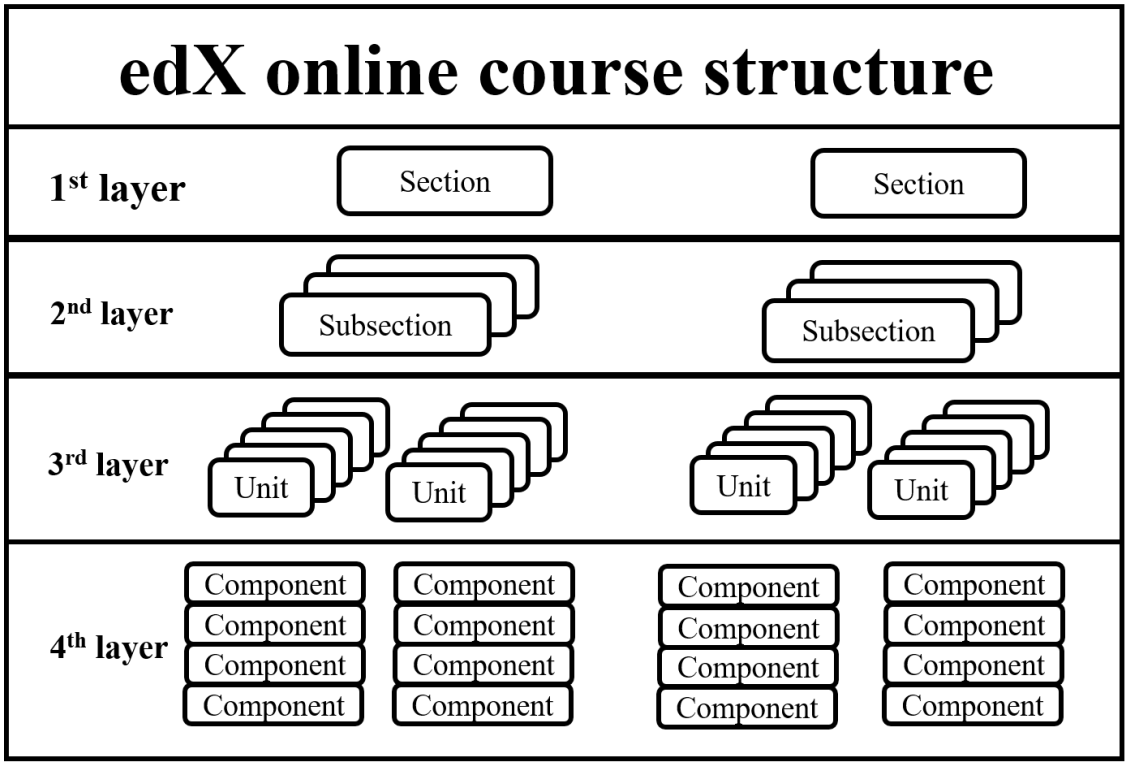
Content Modification Tool for English-to-Japanese Text Replacement

Contents

- edX course structure
- edX Open Learning XML (OLX) structure
- Material preparation
- Content Modification Tool
  - Course structure extraction
  - Component Replacement
- Result

edX Course structure

- edX provides a 4-level hierarchy [1]
- 4th layer component has various type. This tool involves only Text-type component
- For Text replacement, tool only modifies Text component without changing structure of existing course

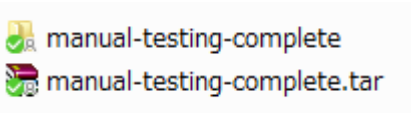


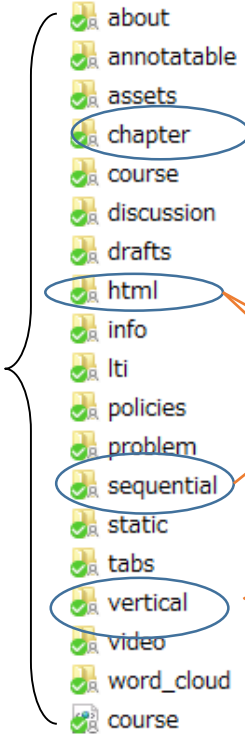
Compare to a course in school
Class, i.e., per week
Topic covered in each class
Chapter of each topic, such as introduction, main content, conclusion,
teaching material , i.e., text document, video

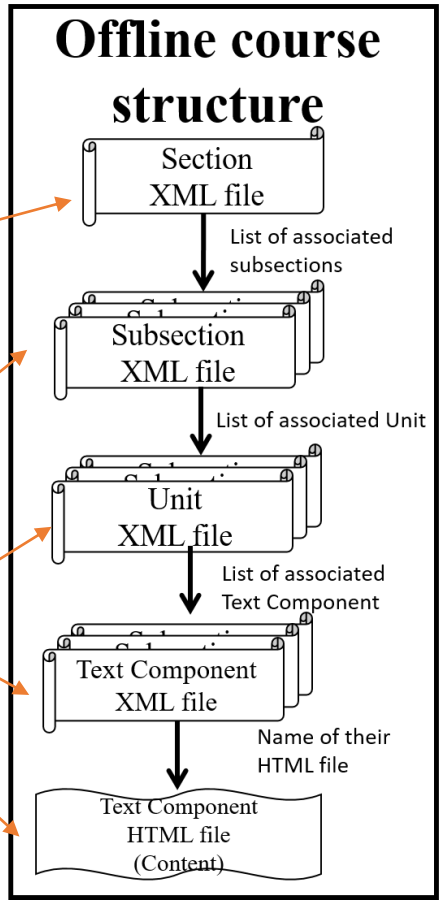
[1] Building and Running an edX course: (2017, April 17). Retrieved from http://edx.readthedocs.io/projects/edx-partner-course-staff/en/latest

edX Open Learning XML (OLX)-1

- When edX course is exported to local computer, data is compressed in tar.gz format [2].
- Content are structured based on OLX as figure below [3]







Section, subsection, unit, and text component information are saved in respective folders (see figure)

[2]edX Open Learning XML Guide: (2017, April 25). Retrieved from http://edx.readthedocs.io/projects/edx-open-learning-xml/en/latest/

[3]edX Manual-testing course: (2017, April 17). https://github.com/edx/edx-platform/tree/master/common/test/data/manual-testing-complete

edX Open Learning XML (OLX)-2

- XML file of these 4-level hierarchy has two information
  - Self-information
  - List of filename of lower layer

Section name

```
<?xml version="1.0"?>
- <chapter display_name="New Section 9 - Zoomable Image">
    <sequential url_name="0cd40e13b4a045aba07d4f626c8b32de"/>
  </chapter>
```

Subsection filename

Subsection name

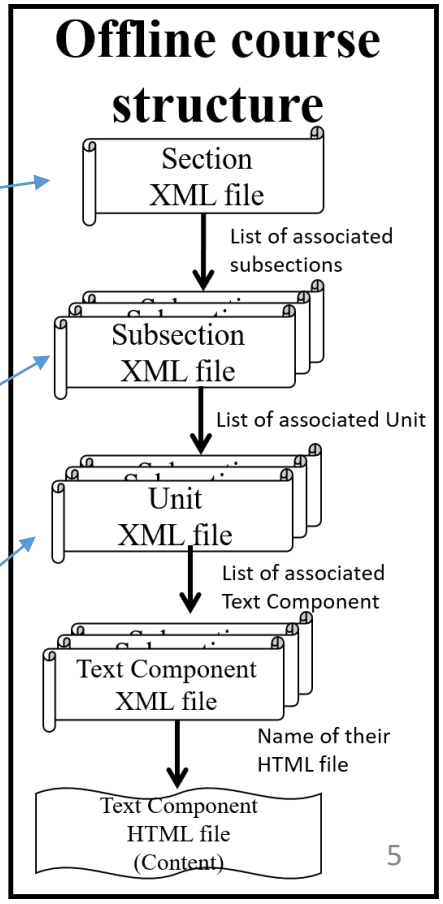
```
<?xml version="1.0"?>
- <sequential display_name="New Subsection 1.3">
    <vertical url_name="c641d8fe821440bea1781287d05fdc82"/>
  </sequential>
```

Unit filename

Unit name

```
<?xml version="1.0"?>
- <vertical display_name="Online Assessment Administration">
    <html url_name="94ae85ee4740475f980d6f838e69b6a1"/>
    <html url_name="dfc3db8f44aa4a3cb563d718e3b30464"/>
    <html url_name="4dbc9090e53845dbbca1d94d348e4d12"/>
  </vertical>
```

Text component filenames



edX Open Learning XML (OLX)-3

- Text component consists of two files
  - XML file

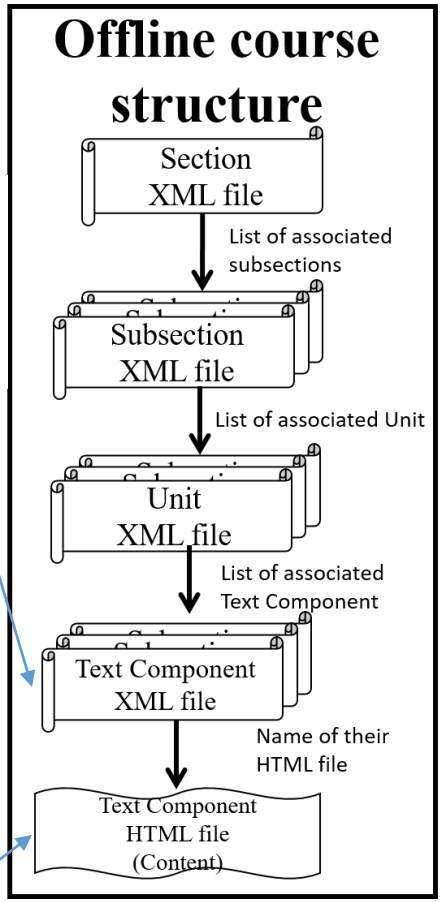
Text component name

```
<?xml version="1.0"?>
<html editor="raw" display_name="Introduction"
    filename="00b99bf259984bd896cb30ee6f8792fd"/>
```

HTML filename

- HTML file
  - plain text content in HTML format

```
<h1>Inquiry-Based Content Presentation Example</h1>
<link href="/static/stylesheet_cocd.css" rel="stylesheet" type="text/css" />
<h2>Introduction</h2>
<p>This sequence of lessons and exercises is an example of inquiry-based content presentation. The
sequence begins with a standard text content presentation style that explains scansion and leads into
an inquiry-based presentation of the concept of feet and specific types of feet (such as anapests and
iambs). Students are asked to perform scansion on the beginning of &amp;ldquo;The Night Before
Christmas&amp;rdquo; so that they can discover on their own that there is a regular pattern before being
introduced to meter in poetry. The target audience is students with an understanding of syllables but
no knowledge of or experience with meter in poetry.</p>
<p>The sequence of lessons begins on the next page.</p>
<p></p>
```



Specifically, tool replaces HTML file of English text with translated text !!!

Material preparation

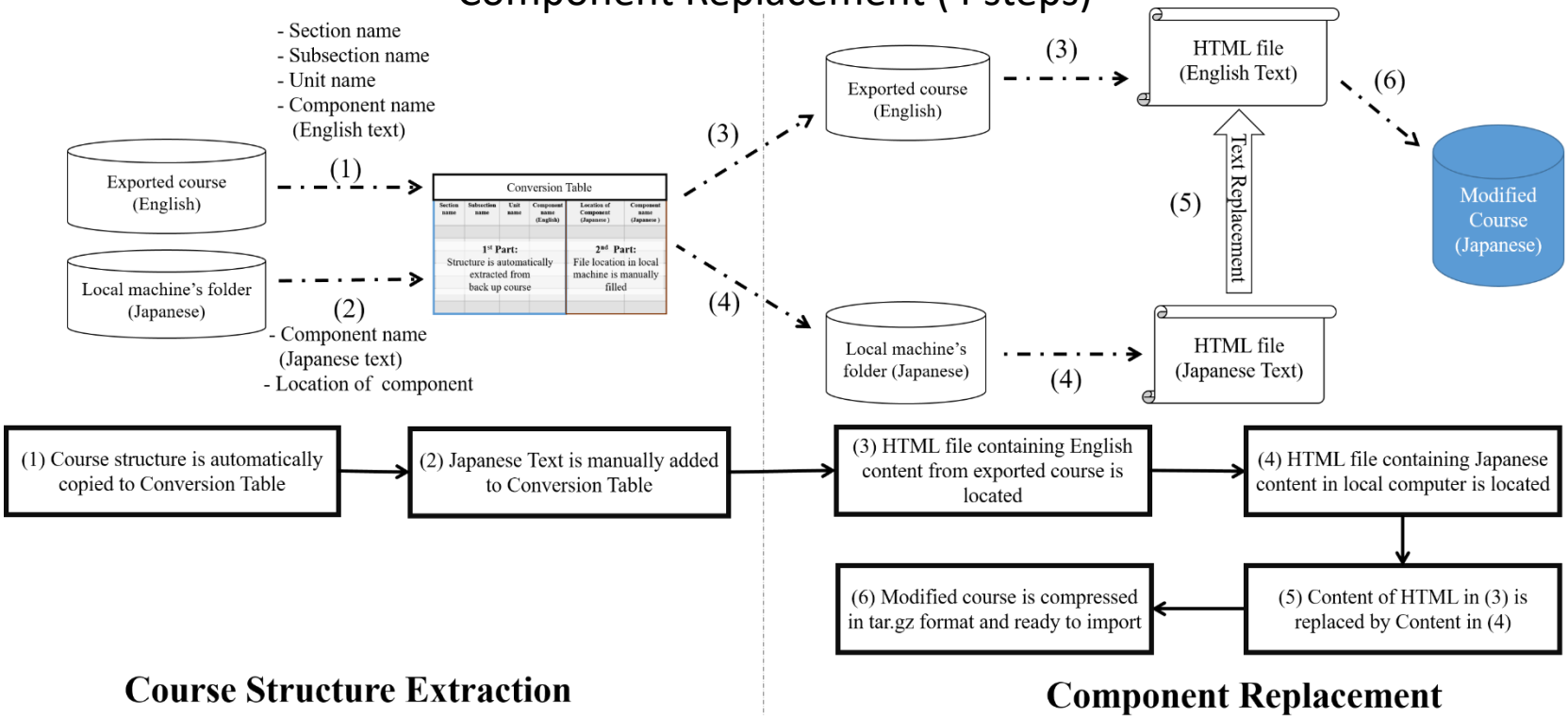
- Modification program.
  - In our work, it is a Python 2.7 script
- Translated material saved in HTML format
  - In our work, translated text (Japanese) was prepared in Microsoft word
  - Then, converted by using '**save as Web page filtered'**

Content Modification Tool

Goal: Replaces HTML file of English text in Exported file with translated text

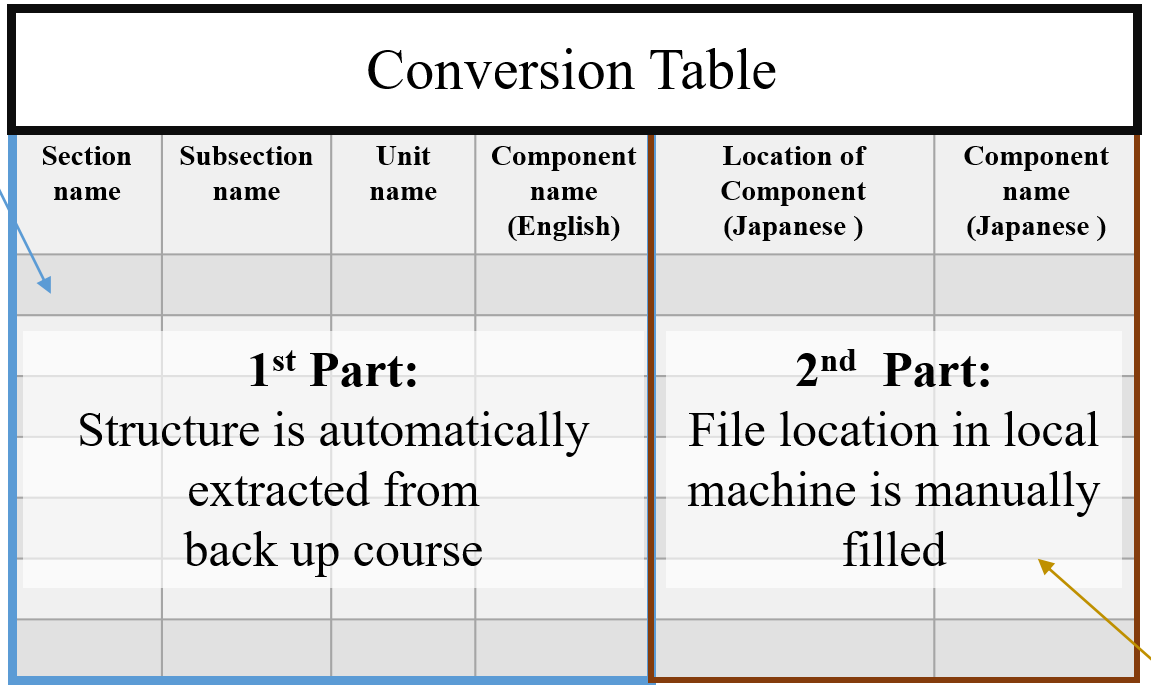
Our tool has two processes

- Course structure Extraction (2 steps)
- Component Replacement (4 steps)



Course structure Extraction

Course structure is automatically copied to Conversion Table
- Python script searches through exported course (as slide 4)
- 4-level hierarchy names are copied into 1st Part of Conversion Table



Japanese text is manually added to Conversion Table
- HTML filename of Japanese text and directory should be manually added in 2nd Part of Conversion Table **according to course structure in 1st part**

Example of Conversion Table

From exported course

From local computer directory

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
163	Content Presentation	Overview	Videos: Location Shooting	Resources	/5) content presentation/1) Overview/unit copy 10	comp4
164	Content Presentation	Overview	Videos: Location Shooting	Pros and Cons	/5) content presentation/1) Overview/unit copy 10	comp5
165	Content Presentation	Overview	Learn By Doing	Learn By Doing	/5) content presentation/1) Overview/unit copy 11	comp1
166	Content Presentation	Overview	Synchronous Virtual Classrooms	Synchronous Virtual Classrooms	/5) content presentation/1) Overview/unit copy 12	comp1
167	Content Presentation	Overview	Synchronous Virtual Classrooms	Description	/5) content presentation/1) Overview/unit copy 12	comp2
168	Content Presentation	Overview	Synchronous Virtual Classrooms	Resources	/5) content presentation/1) Overview/unit copy 12	comp3
169	Content Presentation	Overview	Synchronous Virtual Classrooms	Pros and Cons	/5) content presentation/1) Overview/unit copy 12	comp4

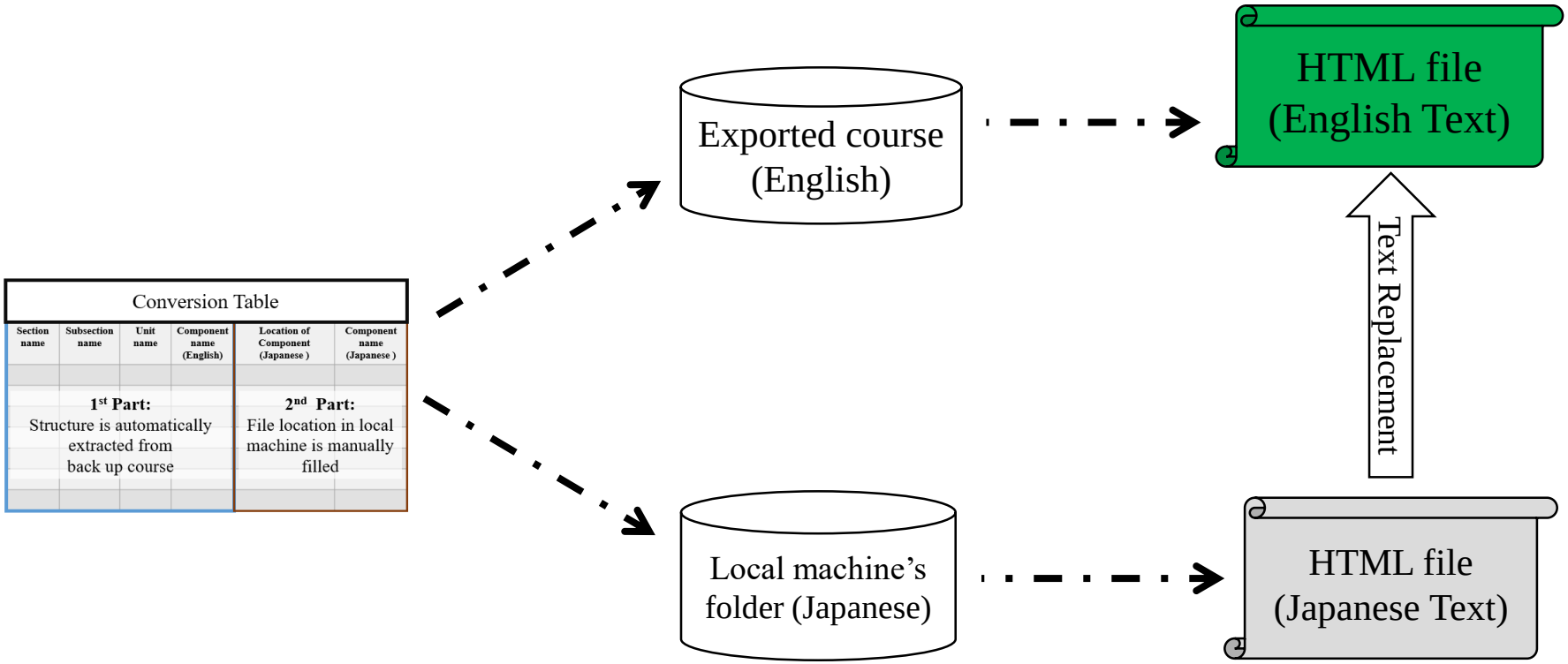
1st Part: automated task

2nd Part: manual task

Microsoft Excel is used for making Conversion table

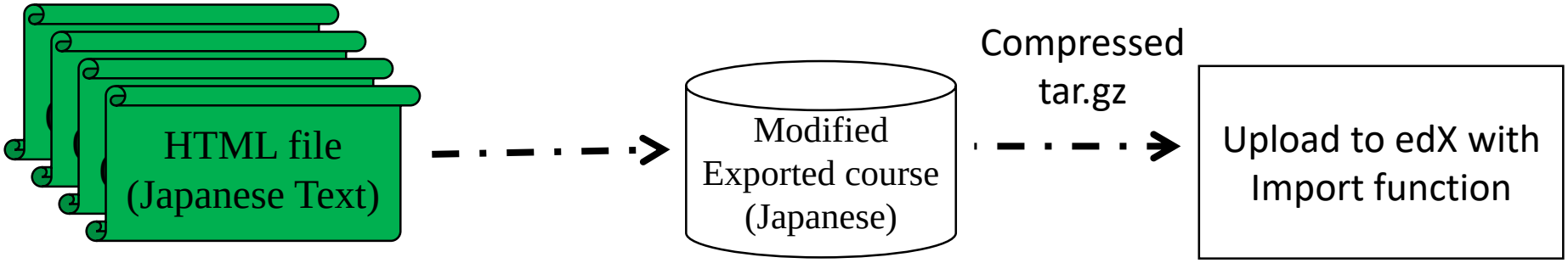
Component replacement - 1

- Tool locates
  - a target HTML file of English text in exported course
  - A designed HTML file of Japanese text in local directory



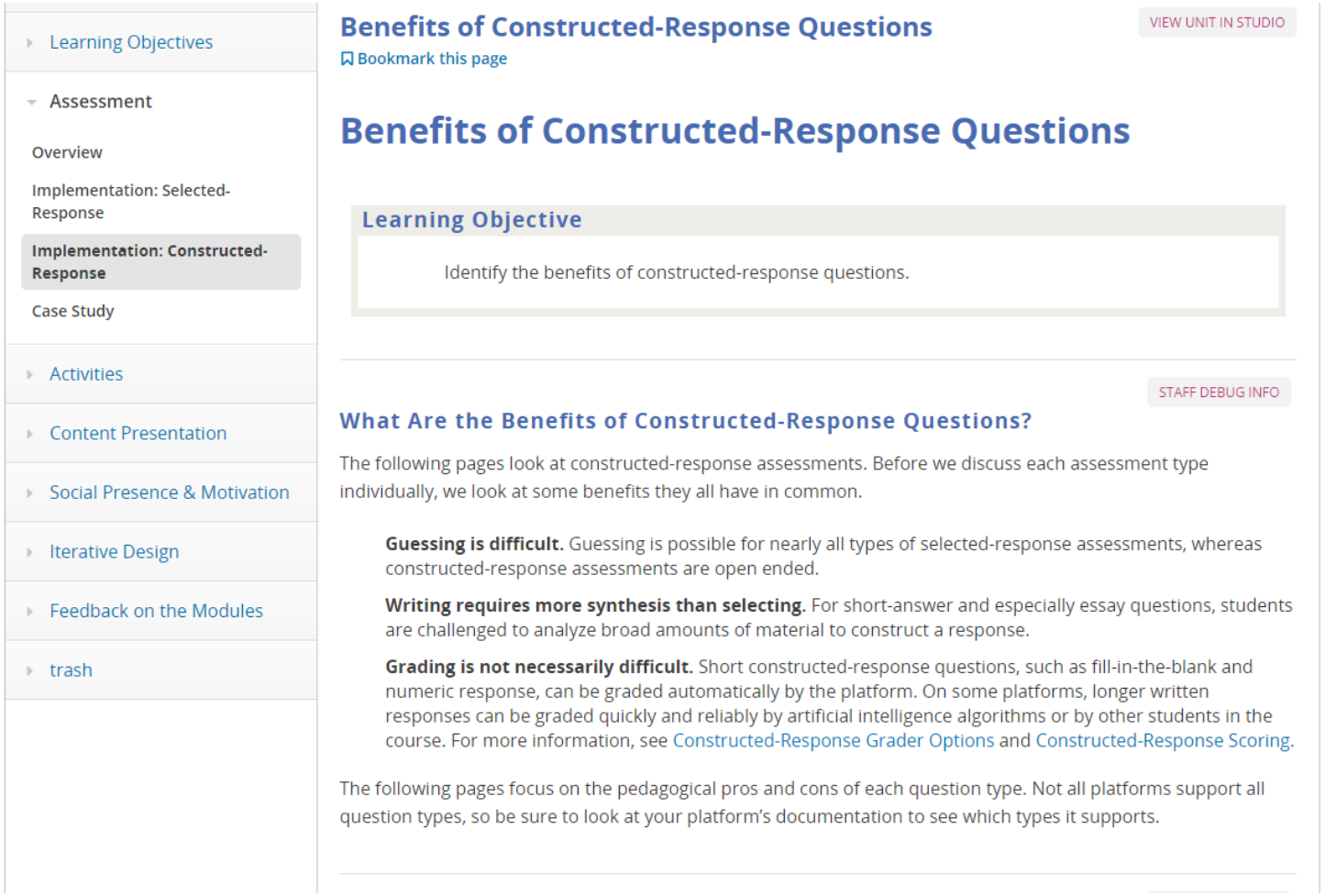
Component replacement - 2

- Tool iteratively replaces all text listed in the Table
- Exported course is now filled with Japanese text
  - Tool compresses the course (slide 4) into tar.gz format
- Manually, Upload the modified course into edX



Result in edX (Before)

Contents are in English



Result in edX (After)

Note that our tool changes only text content. Section,subsection,unit, and text component remains unchanged (English)

Contents are changed to Japanese