**Project details:**

Java Script for TEI2-XML to Word

I'm looking for a skilled Java developer to create a script that can convert TEI2-XML files into Word documents. The main goals of this project are to maintain the original formatting, especially to make the Word documents look as close as possible to the original PDFs.

Key Requirements:

- \*\*Preserving Formatting:\*\* The script should be able to match the original PDFs, maintaining reference, formatting, and layout as much as possible.

- \*\*Chapter Titles and Subtitles:\*\* These should be mapped to Word Styles for consistency and ease of navigation.

- \*\*Manual Adjustments:\*\* While the script should do the bulk of the work, some manual adjustments will be needed. This includes indenting each paragraph, and making necessary adjustments to tables, figures and font styles and sizes.

Ideal Candidate:

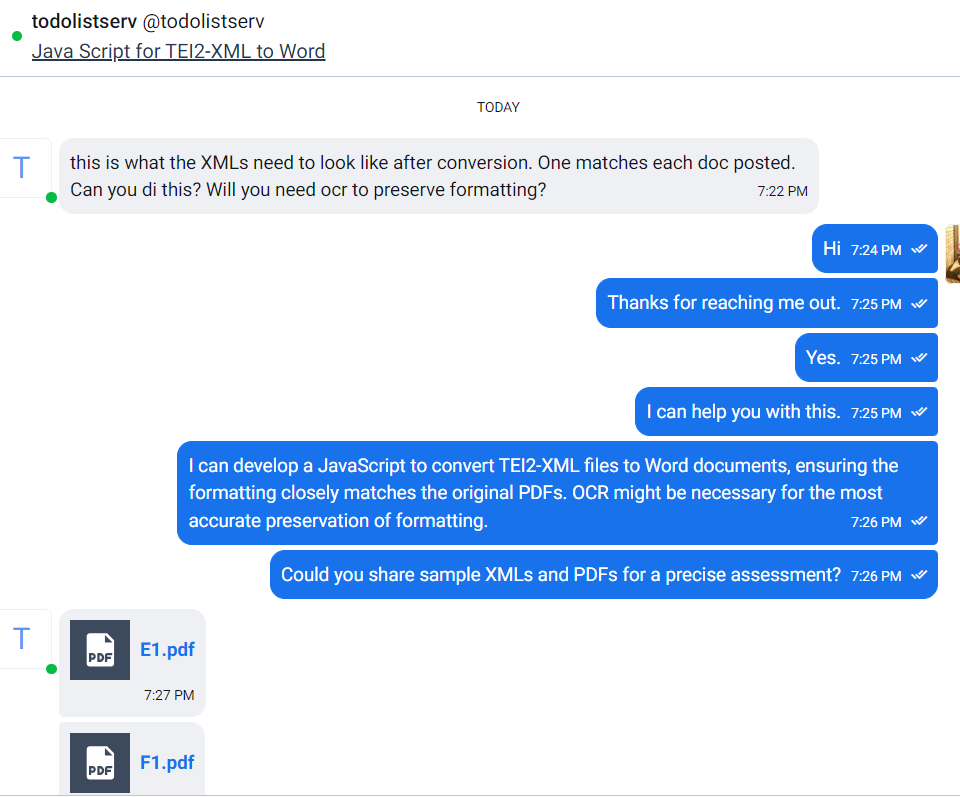
- Proven experience in Java development, previous experience with similar tasks is a plus.

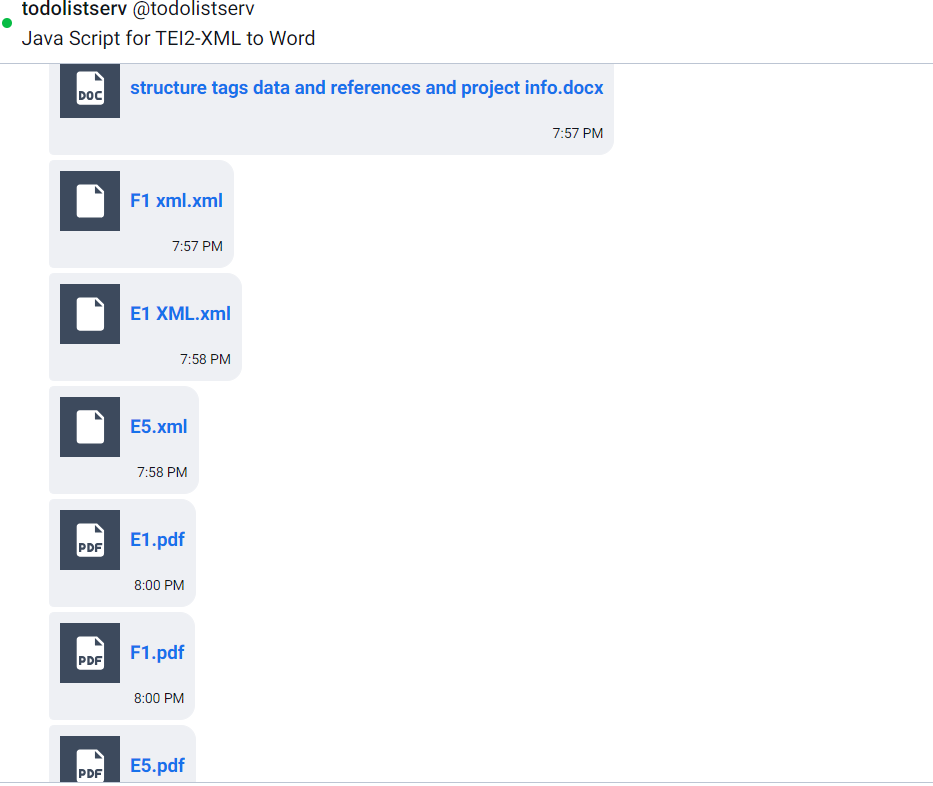
- Familiarity with TEI2-XML and Word documents.

- Attention to detail and the ability to manually adjust documents to mirror the source material.

client message:

This script will run on 7 documents. The Word document will be created from the XML-TEI documents and we will use the PDF versions to check for any issues with formatting. All references MUST be preserved, including indexes. Issues I anticipate: The book indents each line of a paragraph. The XML does not tell us that, so the script will need a manual adjustment to tell it to do that. I think the XML does not give us enough information to recreate tables, so they may need reconstructed manually from PDF. Other than adding word styles for organization and mapping footnotes etc, everything else needs to replicate the PDF as far as formatting. However, the pagination does not need to match because I want the indexes and footnotes to accurately reflect the correct page if I change font sizes etc. It will be in Word 2021, no preference between doc and docx. I have included 2 input and 2 output, and another doc with some general info about reference structure and tags. I may need ongoing support. Ideally I would like the project done in 3-5 days.





These are the files give by the client:  
https://we.tl/t-dFR0RyHkyy