

PDF to EPUB converter

WHAT?

PDF to EPUB converter is a software solution that simplifies the process of turning PDF formatted books to EPUBs that are more suitable for various E-readers. It gives the user wide range of tools to work with. For example tools for extracting paragraphs, headings, footnotes and pictures. After that, the user can add necessary metadata and the EPUB is ready for downloading.

WHY?

Eesti Digiraamatute Selts (Estonian Digital Books Association) was interested in an application that could significantly decrease the time that it took to convert PDF books to EPUBs. In addition, they were mostly outsourcing the conversion to external contractors who had necessary knowledge in HTML, CSS and EPUB specific formatting. Our goal was to make that process less dependent on technical skills and reduce the average amount of time it took to convert a book.

HOW?

Our application's first step was to generate rough estimations on how difficult the conversion process was going to be. This was done mostly by counting different elements from the PDF. After that, the PDF was converted to HTML although it kept its original layout. Then the user has different tools that were done in JavaScript at their disposal. At first they should select the ending point of the chapter. The next step would be to eliminate unnecessary elements like page numbers between paragraphs. After that footnotes can be selected. Then a suitable heading and all the paragraphs. Last step is the pictures. If there are more chapters left the cycle continues, otherwise, metadata can be added and after that, the EPUB file is generated.

WHAT'S NEXT?

PDF to EPUB converter is not a finished product yet. It evolves on a day-to-day bases and there are many more features that it is going to obtain in the future. For example at the moment it converts PDF to EPUB2 standard, but there is a new standard available - EPUB3. Also some other aspects like footnote linking and separate covers are potential features

TEAM Sebastian Värvi (BSc 3y), Karl-Mattias Tepp (BSc 3y), Meelis Tapo (BSc 3y), Silver Vapper (BSc 3y) – all from Faculty of Science and Technology, Institute of computer science, Mentor: Evert Nõlv

<http://pdf-to-epub.weebly.com/>

