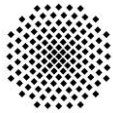


## Instructions: Creating a carcass

Step	Tips and solutions
<b>1. book/ text scanning</b>	<ul style="list-style-type: none"><li>• book/text should have few spots/kink or tears</li><li>• scan with at least 400 dpi and full colour</li><li>• ? that's the attitude that Abbyy used to call her? (for fracture texts) delivers the best results (i. e. the least errors)</li><li>• (→ File: 01-Documentation_Preliminary_work_Fracture_recognition and in detail: Documentation_Abbyyy_Krueger ")</li><li>• (at the UB book scanner you can only decide on the color of the scans. UB book scanner only scans with 300 dpi)</li><li>• if possible, do not scan double pages, but scan each page separately (this is more convenient later when converting with "Abbyy")</li><li>• → tools for cutting the pages are also available</li><li>• (please refer to instructions "02-Instruction AbbyyFineReader_OCR of fracture texts")</li><li>• If possible, scan output that does not contain any line numbering, because it is difficult to cut them off afterwards. In addition, hyphens are not joined in OCRn if there is a number between them.</li></ul>
<b>2. convert book/text to OCR</b>	<ul style="list-style-type: none"><li>• separate instructions for the program "Abbyyy" (incl. instructions how to separate double pages with the help of a tool? s. PDF "Instructions AbbyyFineReader_OCR from Frakturtexten")</li><li>• For comparison, if the conversion with "Abbyy" is worthwhile for a certain book/text, a comparison with "Double-Keying" can be revealing (if the text is in good condition, the conversion with "Abbyy" should go much faster than with the "Double-Keying" method).</li><li>• if necessary, it is also worthwhile to perform the conversion with "Adobe". For a documentation of the comparison "Abbyy vs. Adobe" see the folder "Abbyyy vs. Adobe".</li></ul>



<b>3. collect texts</b> <b>(Alternative to item 1 - 2)</b>	<ul style="list-style-type: none"><li>• different sources are available (TextGrid, The German Text Archive, Gutenberg,...)</li><li>• Textgrid? texts cannot be downloaded in. txt format</li><li>• texts cannot be downloaded, but must be marked in the browser and copied to an editor.</li><li>• Kindle (Amazon) Florian Barth downloads texts here</li></ul>
<b>4. sort carcass</b>	<ul style="list-style-type: none"><li>• There are multiple renaming tools available for this purpose.</li><li>• (separate manual available? PDF: "03-Tutorial Total Commander multiple rename tool")</li><li>• It is also helpful to sort the files into folders, which for example cover 50 years.</li><li>• Create database with Excel[in progress (April 2016)]</li></ul>
<b>5. normalize carcass</b>	<ul style="list-style-type: none"><li>• ensure uniform coding (e. g. UTF-8)</li><li>• ensure uniform spelling</li><li>• Programs for normalization (spelling) of texts:<ul style="list-style-type: none"><li>→ CAB (Anjas Overview)</li><li>→TICCL</li><li>→VARD2</li><li>→CorA</li></ul></li><li>→ written by computer linguist (Sarah Schulz/University of Stuttgart) Documentation of the problems and advantages see files "S. Schulz (Normalization Tool) - Documentation" and "15.16.17.18. century _what was normalized _Sarah Schulz's Tool -Dilan" and also "PROBLEME_Schulz' tool (Table final version)Dilan").</li><li>• further points to follow.....</li><li>•<ul style="list-style-type: none"><li>• Translated with <a href="http://www.DeepL.com/Translator">www.DeepL.com/Translator</a></li></ul></li></ul>