

Lab 11 + 12 - Creating reports

Script Languages

Learning goals

1. Export Word file with mixed content.
2. Prepare custom graphics.

Exercises

1. Preparation

1. Download an html version of a book of your choice from [Project Gutenberg web page/](#). Synchronize your efforts with other students and to choose a unique book. Read carefully the whole assignment and choose a book which could be easily used to solve it.
2. Install hydra package ([hydra.cc](#)) to store and manage configuration of your application.

2. Parsing HTML

1. Develop a function to download an html version of a selected book from Project Gutenberg. Invoke your application using the identifier of a book, e.g.

```
python app10.py 7849
```

2. Extract book title, author's name and the first chapter of the book.
3. Count the number of words in each of paragraphs of the first chapter. Round a number of words to the 10 (e.g. 123, 127, 129 all become 120). Sort the numbers from the shortest to the longest, and count number of paragraphs with the same number of words. Create a plot showing a distribution of lengths of paragraphs in the first chapter.
4. Download first picture (from the book content) from Project Gutenberg website. Link to the picture must be extracted from the text.
5. Crop and resize the picture from the previous task using reasonable factors of your choice. Do not hardcode those factors in the code - the appropriate values should be either stored in a configuration file or given by the user when invoking the application. Embed a watermark containing a caption „Downloaded from Project Gutenberg“ in the image. Place it in the lower left corner. Delete the original file.

3. Creating docx file.

1. Generate a Word document including:
 1. title page with:
 - title of the book,
 - Modifier image from task 2.5
 - author of the book,
 - author of the report (your name). This value should be stored in a configuration file.
 2. Chart pages with plots of distributions of lengths of paragraphs in every chapter. Each plot should be placed on different page of the report. Add a description at the top of such page containing the number of the chapter and its title (if present).
 3. Final page with
 1. A plot presenting number of paragraphs in subsequent chapters of the book.
 2. A summary table with a minimal and maximal number of words in chapters, average number of paragraph in chapters, average number of words in paragraphs etc. Devise your own metrics.
2. The generated docx filename should contain the date and time of file execution and the id of processed book.