**WebApp container requirements**

**Must have**

* Users must be able to upload a text for processing
* Users must be able to use the product without violating any privacy rights

**Should have**

* Users should receive valuable information about how the product processes their texts
* Users should be able to find information about how to format their texts before analysing
* Users should be informed about the scope and contact details regarding the product
* Users should be able to upload multiple files at the same time for processing
* Users should be able to find information about how to interpret their results

**Could have**

* The entire system could compile a zip file of all the result files to avoid cluttering and confusion

**Would be nice to have**

* Would be nice if users could have the option to use the app as an executable to bridge data protection issues

**Machine Learning script requirements**

**Must have**

* The Machine learning script must be able to work with different file extensions (txt, pdf, word etc)
* The Machine learning script must be able to convert raw text data into necessary formats and scores for topic modelling (term frequency, inverse document frequency, tf-idf scores, document-term matrix, bi-grams, n-grams)
* The Machine learning script must be able to run LDA topic modelling on the transformed input files with pre-defined parameters
* The Machine learning script must be able to provide an easily readable and accessible output of the identified topics and the contribution of the most important words

**Should have**

* The Machine learning script should be able to automatically calculate the optimal parameters for the machine learning models of their results
* The Machine learning script should be able to provide an annotated version of the original files by highlighting the sentences which include at least one of the most important words

**Could have**

* The Machine learning script could provide visualization of the most important words and their contribution to the topics
* The machine learning script could return intermediate objects as part of the results (tf, idf, tf-idf scores, document-term matrix) so users are allowed to run their own machine learning scripts on these files.
* The entire system could compile a zip file of all the result files to avoid cluttering and confusion

**Would be nice to have**

* Would be nice for users to have the option to choose from different types of machine learning techniques for text processing

**Non-functional requirements**

* **Availability**: Access to the software mustn’t be limited to any device, location or time.
* **Performance**: The software should be responsive and perform the analysis with as little waiting time as possible
* **Security and Privacy**: The software must try to adhere to GDPR regulation. The handling of the data must be clearly displayed to the user.
* **Extensibility and scalability**: The container should be developed to support the further addition of other machine learning scripts
* **Usability**: Users should be able to easily navigate through the container and they should be able to access and interpret their results with relative ease.