

			HD	D	C	P	F
Task 1 (60%)							
	Output (50%)						
		Correct structure for the CSV file as specified in the sample.csv	It is a must. Note that the output files should be parsable (i.e. satisfy all the requirements of the CSV files) and also be in the pre-defined structure. Unparsable outputs and the output files that have any deviations from the sample output, as specified in the specs, will receive zero marks for the output.				
		Review data	Less than 4% of entries are incorrect.	Less than 8% of entries are incorrect	Less than 15% of entries are incorrect.	Less than 30% of entries are incorrect.	More than 30% of entries are incorrect.
	Methodology (25%)		The report has demonstrated a good solution in a proper way. The methodology consists of all required steps (e.g., using correct API functions for input and output, developing correct regular expression, retrieving correct information, etc) for the tasks, and producing the correct output.	The report has demonstrated a reasonable solution. One key step (e.g., using correct API functions for input and output, developing correct regular expression, retrieving correct information, etc) is missing or leads to the incorrect output.	The report has provided a fair solution. Two key steps (e.g., using correct API functions for input and output, developing correct regular expression, retrieving correct information, etc) are missing or lead to incorrect output.	The report has provided a poor solution. Three key steps (e.g., using API correct functions for input and output, developing correct regular expression, retrieving correct information, etc) are missing or lead to incorrect output.	The report has provided a bad solution. Only one or two steps are demonstrated, which clearly cannot achieve the correct output.
	Documentation (25%)		The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained properly and the code is well commented.	The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained fairly and the code is fairly commented. The report can be improved.	The report has some sectioning but it is not well organised. The explanation of the report is limited and the code is well commented.	The report has some sectioning but it is not well organised. The explanation of the report is poor and the code is well commented.	The report has poor/no sectioning and it is not well organised. The explanation of the report is poor and the code is poorly commented.
Task 2 (40%)							
	Output (50%)						
		Correct structure for the output files as specified in the spec	It is a must. Note that the output files should be parsable and also be in the pre-defined structure (i.e., having correct delimiters and datatypes). Unparsable outputs and the output files that have any deviations from the specification, will receive zero marks for the output.				
		Vocab (35%)	Less than 3% of the vocab entries are wrong.	Less than 6% of the vocab entries are wrong.	Less than 10% of the vocab entries are wrong.	Less than 15% of the vocab entries are wrong.	More than 25% of the vocab entries are wrong.
		Sparse file (15%)	Less than 1% of the vocab entries are wrong.	Less than 3% of the vocab entries are wrong.	Less than 10% of the vocab entries are wrong.	Less than 15% of the vocab entries are wrong.	More than 25% of the vocab entries are wrong.
	Methodology (25%)		The report has demonstrated a good solution in a proper way. The methodology	The report has demonstrated a reasonable solution. One key step (e.g., pdf	The report has provided a fair solution. Two key steps (e.g., pdf	The report has provided a poor solution. Three key steps (e.g.,	The report has provided a bad solution. Only one or two steps are

			consists of all required steps (e.g., pdf parsing, tokenization, stopping, Stemming, etc) for the tasks, and produces the correct output.	parsing, tokenization, stopping, Stemming, etc) is missing or leads to the incorrect output.	parsing, tokenization, stopping, Stemming, etc) are missing or lead to incorrect output.	pdf parsing, tokenization, stopping, Stemming, etc) are missing or lead to incorrect output.	demonstrated, which clearly cannot achieve the correct output.
	Documentation (25%)		The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained properly and the code is well commented.	The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained fairly and the code is fairly commented. The report can be improved.	The report has some sectioning but it is not well organised. The explanation of the report is limited and the code is well commented.	The report has some sectioning but it is not well organised. The explanation of the report is poor and the code is well commented.	The report has poor/no sectioning and it is not well organised. The explanation of the report is poor and the code is poorly commented.