

## The Rubric for A2

		HD	D	C	P	F
	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.				
Output (60%)	Vocab (35%)	80% or more of the vocab entries are correct.	70% - 79% of the vocab entries are correct.	60% - 69% of the vocab entries are correct.	50% - 59% of the vocab entries are correct.	0% - 49% of the vocab entries are correct.
	Sparse file (15%)	80% or more of the count vectors for common vocab are correct.	70% - 79% of the count vectors for common vocab are correct.	60% - 69% of the count vectors for common vocab are correct.	50% - 59% of the count vectors for common vocab are correct.	0% - 49% of the count vectors for common vocab are correct.
	Statistics (10%)	80% or more of the stats are correct.	70% - 79% of the stats are correct.	60% - 69% of the stats are correct.	50% - 59% of the stats are correct.	0% - 49% of the stats are correct.
Methodology (25%)		The report has demonstrated a good solution in a proper way. The methodology consists of all required steps (e.g., downloading files, pdf parsing, tokenization, stopping, Stemming, etc) for the tasks, and produce the correct output.	The report has demonstrated a reasonable solution. One key step (e.g., pdf parsing, tokenization, stopping, Stemming, etc) is missing or leads to the incorrect output.	The report has provided a fair solution. Two key steps (e.g., pdf parsing, tokenization, stopping, Stemming, etc) are missing or lead to incorrect output.	The report has provided a poor solution. Three key steps (e.g., pdf parsing, tokenization, stopping, Stemming, 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 (15%)		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 organized. The explanation of the report is limited and the code is well commented.	The report has some sectioning but it is not well organized. The explanation of the report is poor and the code is well commented.	The report has poor/no sectioning and it is not well organized. The explanation of the report is poor and the code is poorly commented.

Note 1: Both additional terms and missing terms in the vocab are considered as errors.

Note 2: Only the common vocab (i.e., the intersect between the submitted file and the true vocabulary) is going to be considered when evaluating the sparse file.

### Individual Contribution Score (ICS):

The above rubric will be prorated based on each student's ICS, which will be calculated based on the contributions sheet, reflective diaries, and the interview performance (optional). For example: if a student achieves 50% ICS score, he/she will get only half the mark resulting from above rubric. Hence, please document your individual contributions thoroughly and be very well prepared during the interview to explain the steps you have taken to obtain your results based on a true understanding of the underlying concepts.