

A step-by-step approach to solving the problem

1- data pre-processing: I remove all special characters, lowercase the letter, and delete the unnecessary white space.

2- I used the PhraseMatcher routine to locate the technical skill we want to label as hard skill.

3- I used the utility function offset to extract skill position (start index and end index) in each raw skill data and only saved those with matched hard skills. The data saves have technical skill as the first element, starting and ending index of hard skill to extract inside in the specific line as the second element, and the last element contains the label.

5- Data saves contain filter data with the skill we are looking for.

4 - I then extracted unique technical skills matching jargon-filled technical skills.