# Task 1

Overarching issue: extracting required skills from job posting for job seekers

Webcrawler: webcrawler targets domains with job postings

In an ever-competitive job market, finding employment can be quite difficult. Finding a suitable employment opportunity can become quite time consuming for job seekers with a medium sized posting having an average of 450 words [[source]](https://www.linkedin.com/business/talent/blog/talent-acquisition/stats-that-will-change-the-way-you-write-job-posts). A job seeker must read every job posting to find whether they have the required skillset and whether the day-to-day tasks they will be completing are something they are interested in.

A web crawler will be able to target domains that provide job post hosting and extract the textual content from these websites of interest.

The first NLP task will be the summarisation of the job posting through the use of extraction-based summarisation.

The second NLP task will be skill keyword extraction through the use of LSTM and word embeddings, combined with a 2000 row manually labelled training set. The training set is comprised noun-phrases extracted from the first job posting from each sub-classification.