

Project Update: Jobs to Skills

15 November 2021

Nguyen Hoang Giang (HG)	gianghn2@illinois.edu	Captain
Nguyen Thu Giang (TG)	tgn3@illinois.edu	Member

Progress

Please refer to the following table to see the completed and pending tasks regarding this project.

No	Task	Hours allocated	Member in-charge	Progress
1	Crawl job description	5	HG, TG	Completed
2	Dataset preparation and cleaning	5	HG, TG	Completed
3	Topic modelling	10	HG	Starting on 15 Nov
	a. LDA	(5)		
	b. Plsa	(5)		
4	Topic clustering	10	TG	Starting on 15 Nov
5	Analysis of result	5	HG, TG	Starting on 21 Nov
6	Report writing	5	HG, TG	Starting on 21 Nov
Total hours		40		

Challenges

- Crawling of job descriptions has a limited number of API calls (only 1000 per day), hence we need to deploy divide-and-conquer strategy to mine a number of job descriptions each day.

- Job description texts require intensive cleaning due to many non-English symbols.