

LT Resources Project Diary

Ongoing: Reading relevant (and occasionally irrelevant) literature, thinking about project possibilities.

24/11 Downloaded Flickr8k dataset, started exploring data.

27/11 – 7/12 Explored data, tried out different RegEx to capture the right sentences, got counts for various ethnicity words in Flickr8k, chose what RegEx to use.

7/12 Meeting with Dana about project.

8/12 Created small dataset from Flickr8k data (= filtered captions based on RegEx, noted lack of captions mentioning white people, manually inspected and sorted more images to balance dataset).

15/12 Started expanding dataset (more manual sorting).

18/12 Downloaded and started exploring the MS COCO dataset.

21/12 Created additional testset from the MS COCO dataset.

26/12 Read relevant literature, planned continued work, thought about report content and angle.

27/12 Put Flickr8k and MS COCO datasets together, cleaned up code, organized repository.

28/12 Counted expression occurrences, started outlining report.

1/1 Made presentation slides, outlined report, wrote method section.

2/1 Documented code, wrote usage instructions on github.

4/1 Continued writing report.

5/1 Finished writing report.