

### Số lần xuất hiện của các từ

	Tổng số xuất hiện của data	Tổng số xuất hiện	Tổng số từ	Độ phủ
Tổng số từ	441633	441633	3558	100%
Số từ freq >= 2	441633	440252	2177	99.69%
Số từ freq >= 3	441633	439336	1719	99.48%
Số từ freq >= 4	441633	438703	1508	99.34%
Số từ freq >= 5	441633	438043	1343	99.19%

### Top 10 tần suất cao nhất của N-gram

STT	unigram	số lần	bigram	số lần	trigram	số lần
1	<string>	69284	cd <string>	12628	git commit -m	3017
2	ls	30112	cd <directory>	5478	commit -m <string>	2991
3	git	27033	cd ..	5357	git add .	2127
4	cd	25665	vim <file>	5047	-o -name <file>	1757
5	<file>	20754	git add	4298	-name <file> -o	1754
6	sudo	15983	git commit	4063	<file> -o -name	1754
7	<directory>	13924	vi <file>	3493	git add <string>	1709
8	clear	5767	git push	3438	rm -rf <string>	1591
9	vim	5718	-m <string>	3370	git clone <URL>	1443
10	add	5467	git status	3040	sed -i <directory>	1250

### Thống kê N-gram

	Tổng số xuất hiện của data	Tổng số gram	Số từ xuất hiện 1 lần	Tỉ lệ xuất hiện 1 lần
unigram	441633	3558	1381	0.31%
bigram	259374	9683	4260	1.64%
trigram	123926	10197	5026	4.06%

#### Đánh giá mô hình

##### - Gợi ý các từ trên 1 câu lệnh:

- + Thời gian chạy: 0.002704 s
- + Độ chính xác:
  - Top-1 Accuracy: 67.20
  - Top-3 Accuracy: 88.54
  - Top-5 Accuracy: 93.76

##### - Gợi ý câu lệnh tiếp theo:

- + Thời gian chạy: 0.003
- + Độ chính xác:
  - Đúng cả câu trong top 5: 0.4613
  - Các câu gợi ý chỉ cần giống 2 từ đầu tiên của câu lệnh mong muốn
    - Accuracy Top-1: 0.3125
    - Accuracy Top-3: 0.5173
    - Accuracy Top-5: 0.5760