# nl2bash

## Final Results and Discussion

Oisin Doherty (oisind),
Abhinav Gottumukkala (anak4569),
Hans Jorgensen (thehans),
Lauren Martini (lmartini)

# Motivation

- Bash is the de facto language for interacting with the Linux OS
- Bash's power comes from modularity, but man pages only explain individual commands
- The commands needed for complex operations (ex: xargs) can be unintuitive

Example:

**English:** "Change tabs to spaces in every .java file in the current directory"

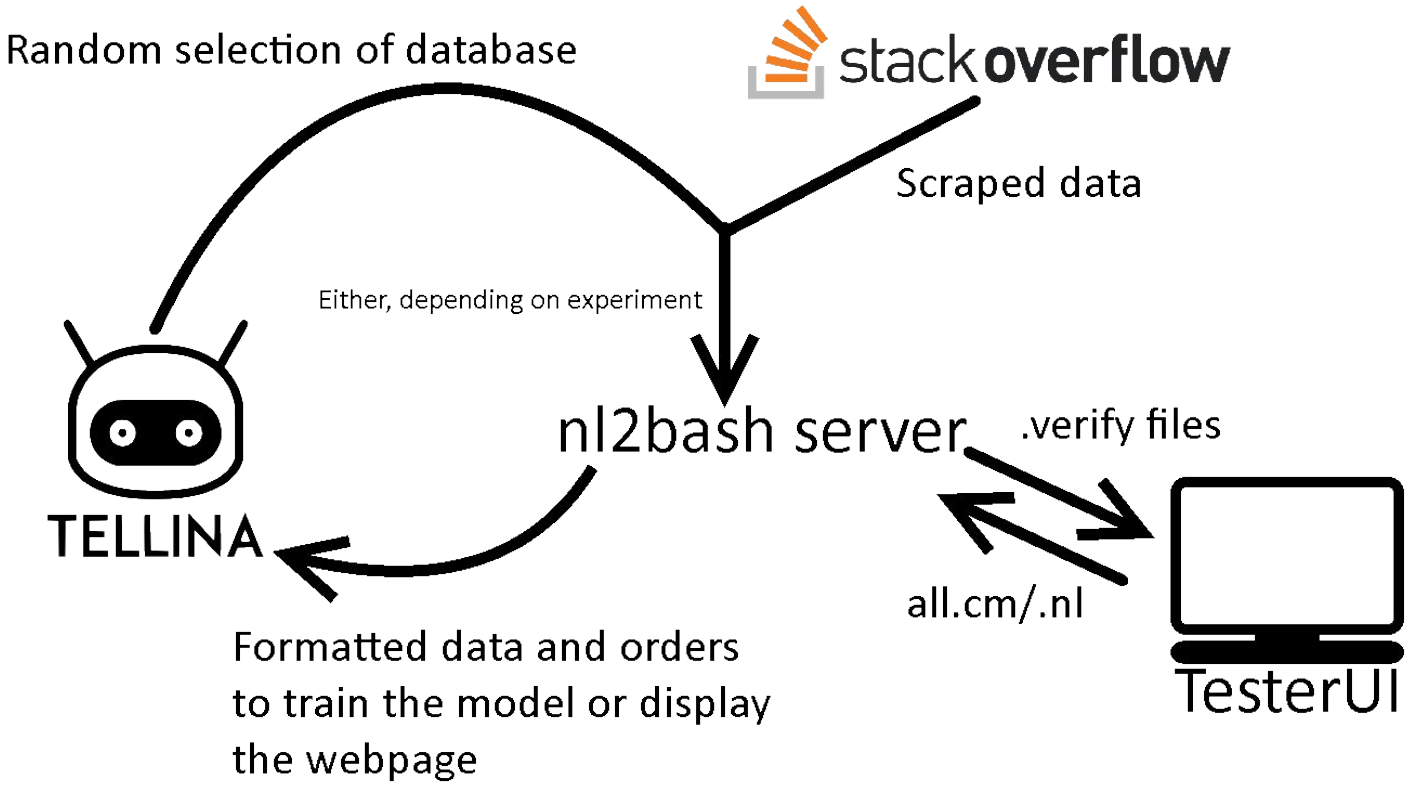
**Bash:**

```
find . -name '*.java' ! -type d -exec bash -c 'expand -t 4 "$0" > /tmp/e && mv /tmp/e "$0"' {} \;
```

# Goals

- Main goal: **Improve the accuracy of Tellina**
- Direction:
  - **Expand Tellina's dataset**
    - Scrape resources such as StackOverflow for English/Bash command pairs
    - Clean and verify this data - crowdsourcing
  - **Improve Tellina's dataset**
    - Develop better cleaning and verification tools to use on the existing dataset - crowdsourcing

# Architecture

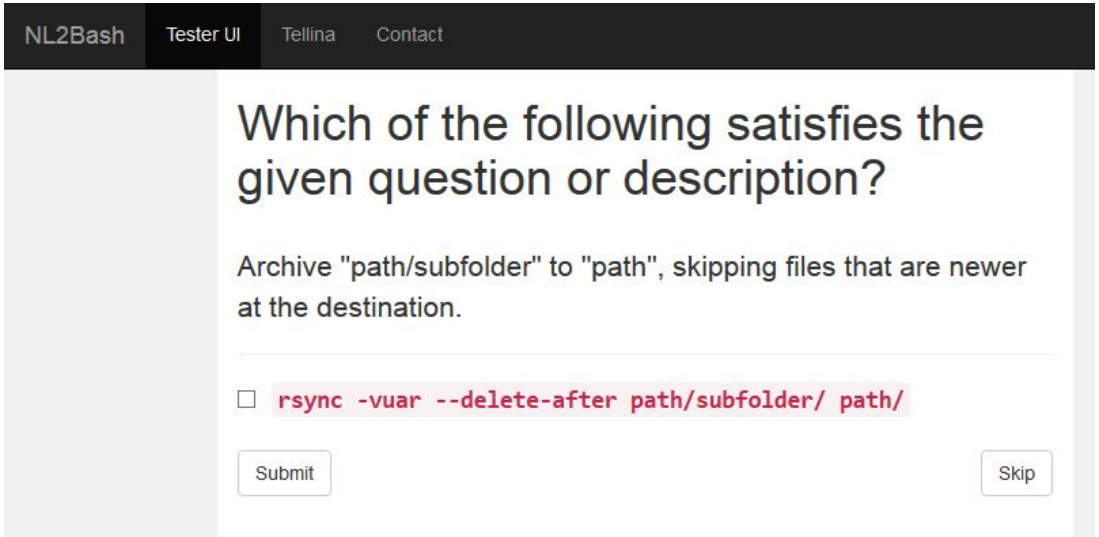


# Stack Overflow Scraping

- Autonomous process to produce data for TesterUI: once activated, no further input is required
- Reads Stack Overflow posts for posts tagged “bash”, saving question and code segments for potential code pairs
  - Can match multiple commands to a description
- Chose to scrape HTML pages instead of using database dumps because the dumps didn’t have any clear means to find code snippets (tags, markdown indicators)

# TesterUI

- Minimal interface for a tester to verify that a command fits its description
- Mechanical Turks task or volunteering (we did it by hand ourselves for our experiments)
- Crowdsourcing or volunteering isn't as accountable, but it's a tradeoff with cost



# Evaluation Methodology

- Metric: Translation accuracy - percentage of English/Bash pairs marked as correct (automatic and manual analysis)
  - Check this for every modified version of the dataset
  - Multiple automatic methodologies (BLEU, TM)
- Goal: More accurate translation than the current version of Tellina

# Experiments

- **Control**: Simply test the original dataset again
- **Cleaned**: Sample the original dataset and remove any false matches we find
  - 423 commands sampled from the original
  - 22% were incorrect matches
- **Augmented**: Add additional scraped and tested pairs from StackOverflow to the dataset (200 pages scraped)
  - 793 potential commands from those pages
  - 9.3% of the matches were good
- **Cleaned & Augmented**: Both clean and augment the dataset as above

| Experiment | Added | Removed | Total |
|---|---|---|---|
| Control | 0 | 0 | 12606 |
| Cleaned | 0 | 92 | 12514 |
| Augmented | 74 | 0 | 12680 |
| Cleaned & Augmented | 74 | 92 | 12588 |

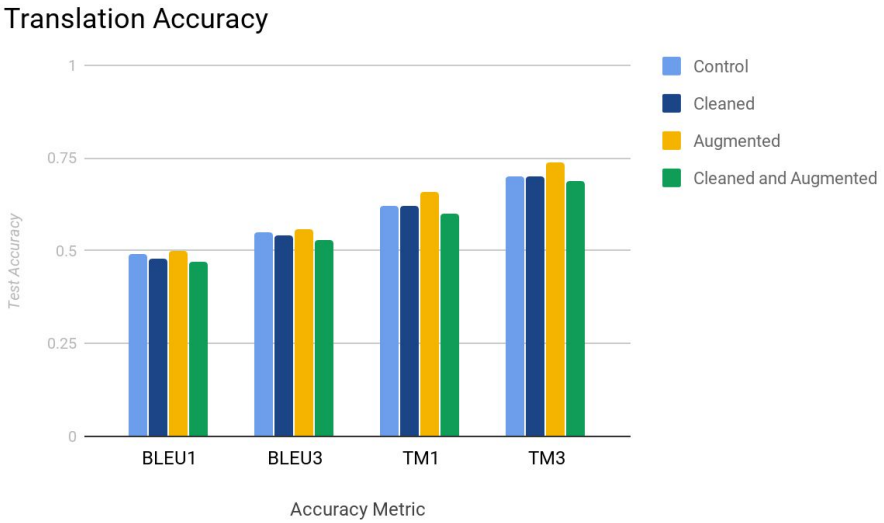
# Results

| Dataset | BLEU1 | BLEU3 | TM1 | TM3 |
|---|---|---|---|---|
| Control | 0.49 | 0.55 | 0.62 | 0.70 |
| Cleaned | 0.48 | 0.54 | 0.62 | 0.70 |
| Augmented | **0.50** | **0.56** | **0.66** | **0.74** |
| Cleaned & Augmented | 0.47 | 0.53 | 0.60 | 0.69 |

BLEU1/3: Bilingual Evaluation Understudy (top candidate vs. top 3 candidates)
TM1/3: Max percentage of close-vocabulary overlap (top candidate vs. top 3 candidates)

# Discussion of Results

- Cleaning the data made it worse in both cases
- Augmenting the data alone was better than the base dataset
- Results suggest more and more varied data improves quality the best



# Next Steps

- Better heuristics for scraping
  - 9.3% of scraped commands verified usable, can be improved
- Continue to gather additional, varied data for further machine translation
- Consider other language training models
  - Word embeddings, other input languages

# References


1. Lin, X. V., Wang, C., Zettlemoyer, L., & Ernst, M. D. (2018). NL2Bash: A Corpus and Semantic Parser for Natural Language Interface to the Linux Operating System. arXiv preprint arXiv:1802.08979.
2. Lin, X. V., Wang, C., Pang, D., Vu, K., & Ernst, M. D. (2017). *Program synthesis from natural language using recurrent neural networks* (Vol. 2). Technical Report UW-CSE-17-03-01, University of Washington Department of Computer Science and Engineering, Seattle, WA, USA.