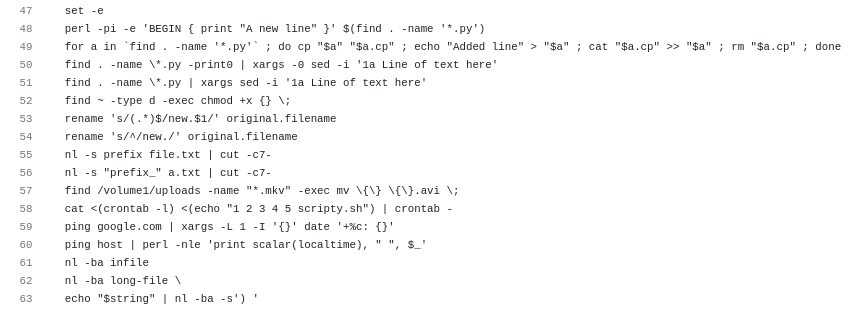
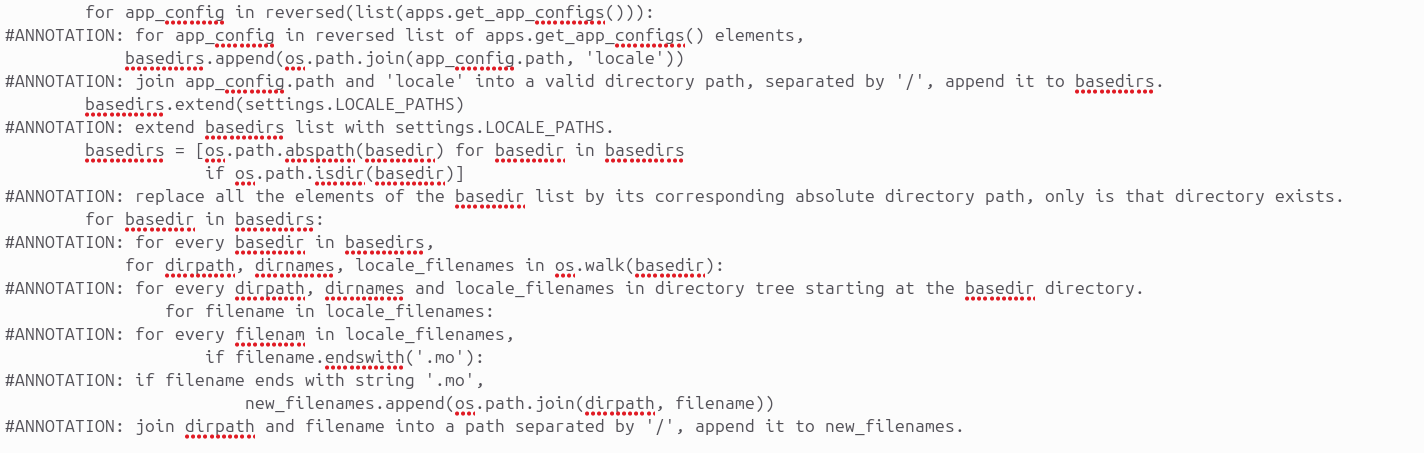
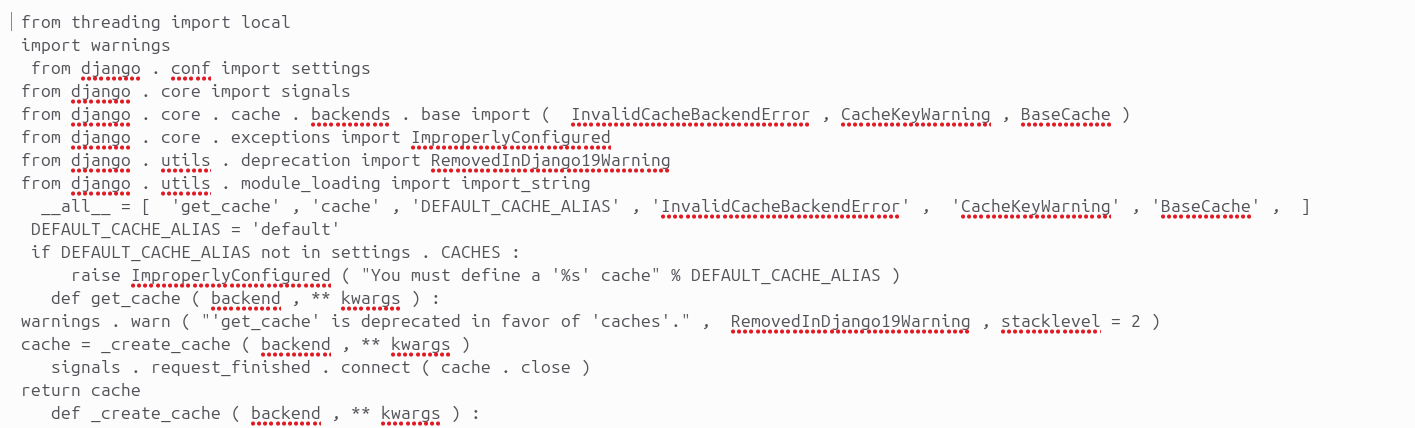
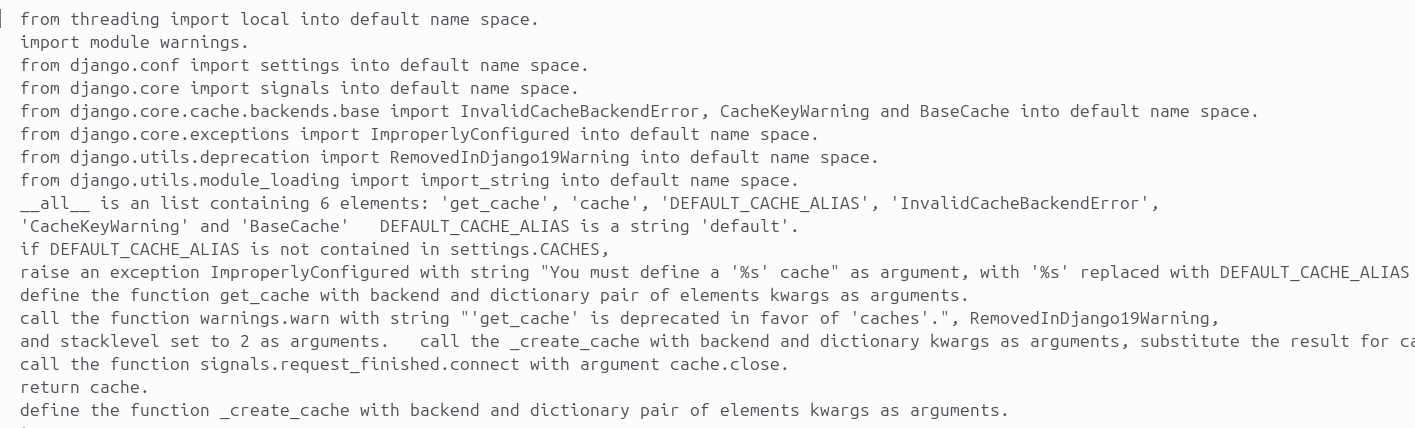
### **Some existing datasets from related tasks:**

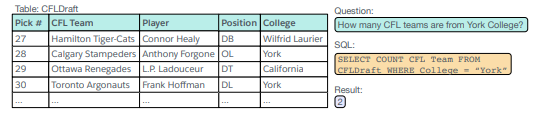
* **NL2Bash contains 12k one-line Linux shell commands used in practice and their natural language descriptions provided by experts. The dataset includes 100+ commonly used shell utilities, providing a rich training and test bed for automatically translating natural language into command lines.**
  + Paper link : <https://arxiv.org/pdf/1802.08979.pdf> .
  + Github repo link : <https://github.com/TellinaTool/nl2bash/tree/master/data> .

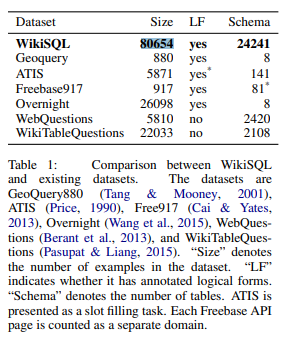


* **The paper demonstrates a tool to generate Pseudo-code from Source Code in Python. They have made their training data available which is basically a django based website where for each line of code there is pseudo code/comment**
  + Paper link: <https://www.phontron.com/paper/fudaba15asedemo.pdf>
  + Dataset link : <https://ahclab.naist.jp/pseudogen/en-django.tar.gz> 

**The paper aims to generate a SQL query from a natural language question and table schema.**

* + Paper Link: <https://arxiv.org/pdf/1709.00103.pdf>
  + Github Repo Link : <https://github.com/salesforce/WikiSQL>





**Spoc Dataset : The dataset contains 18,356 programs for 677 programming problems(easiest level of problems on** [**codeforces.com**](http://codeforces.com/)**). On average, each program has 14.7 lines, and each pseudocode line has about 7.86 tokens. The dataset contains pseudocode,code and testcases to check code correctness.**

Paper/Dataset Link: <https://sumith1896.github.io/spoc/> 