

ClassCorrect

Fixing classification errors

Ivan Teo

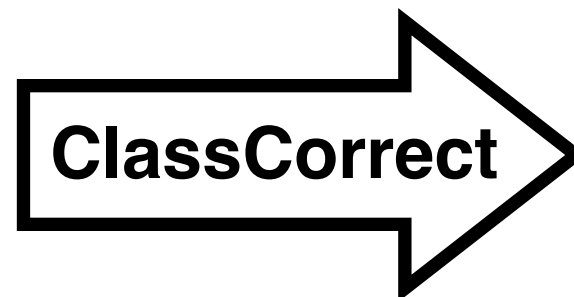


Problem Statement

- Around 30 million entries
- Unstructured descriptions
- About 120 pre-set classes assigned by 3rd party
- Errors in assignment of classes
- e.g. Triage of service requests.

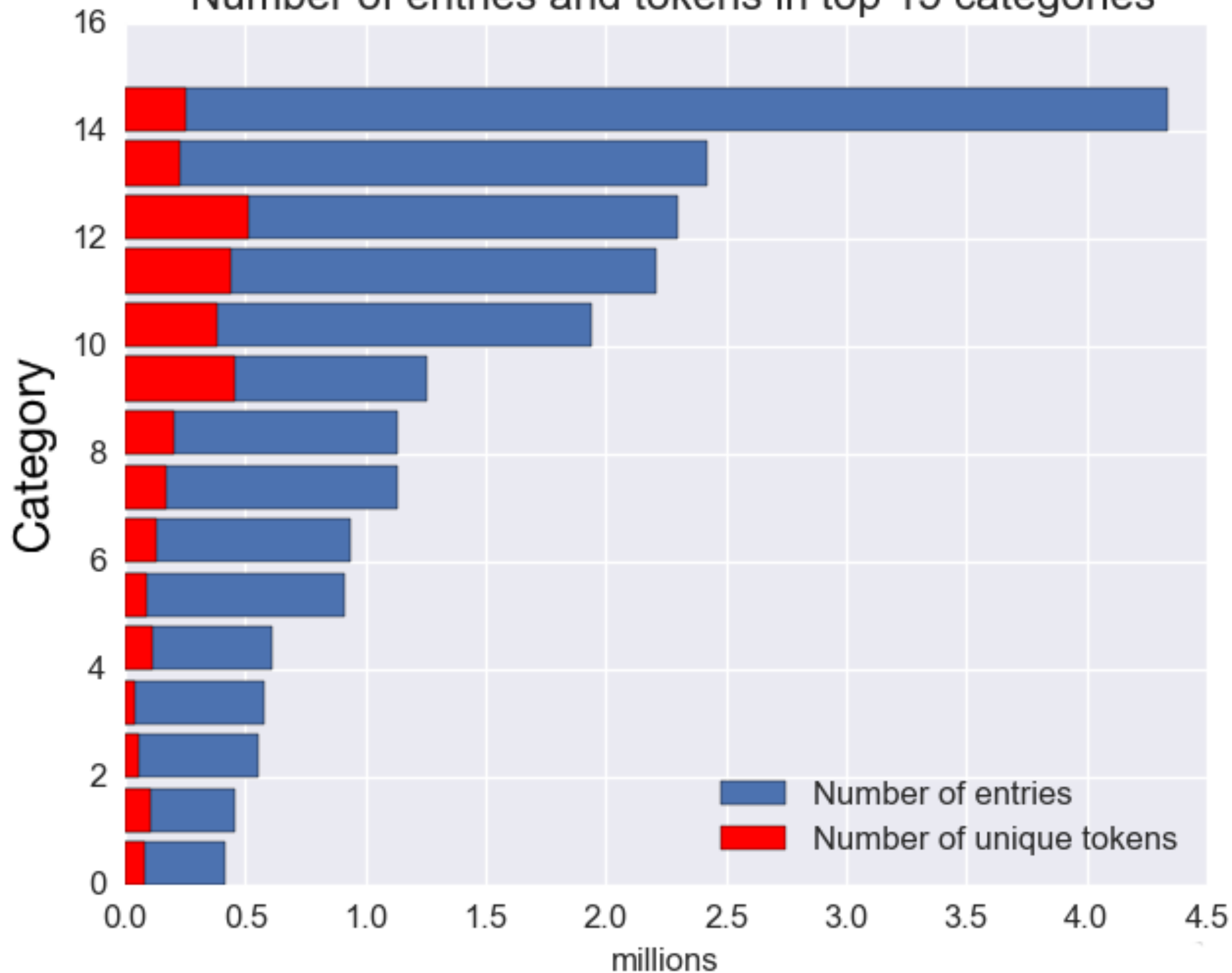
Water leaking from A/C pipe

Class: Plumbing

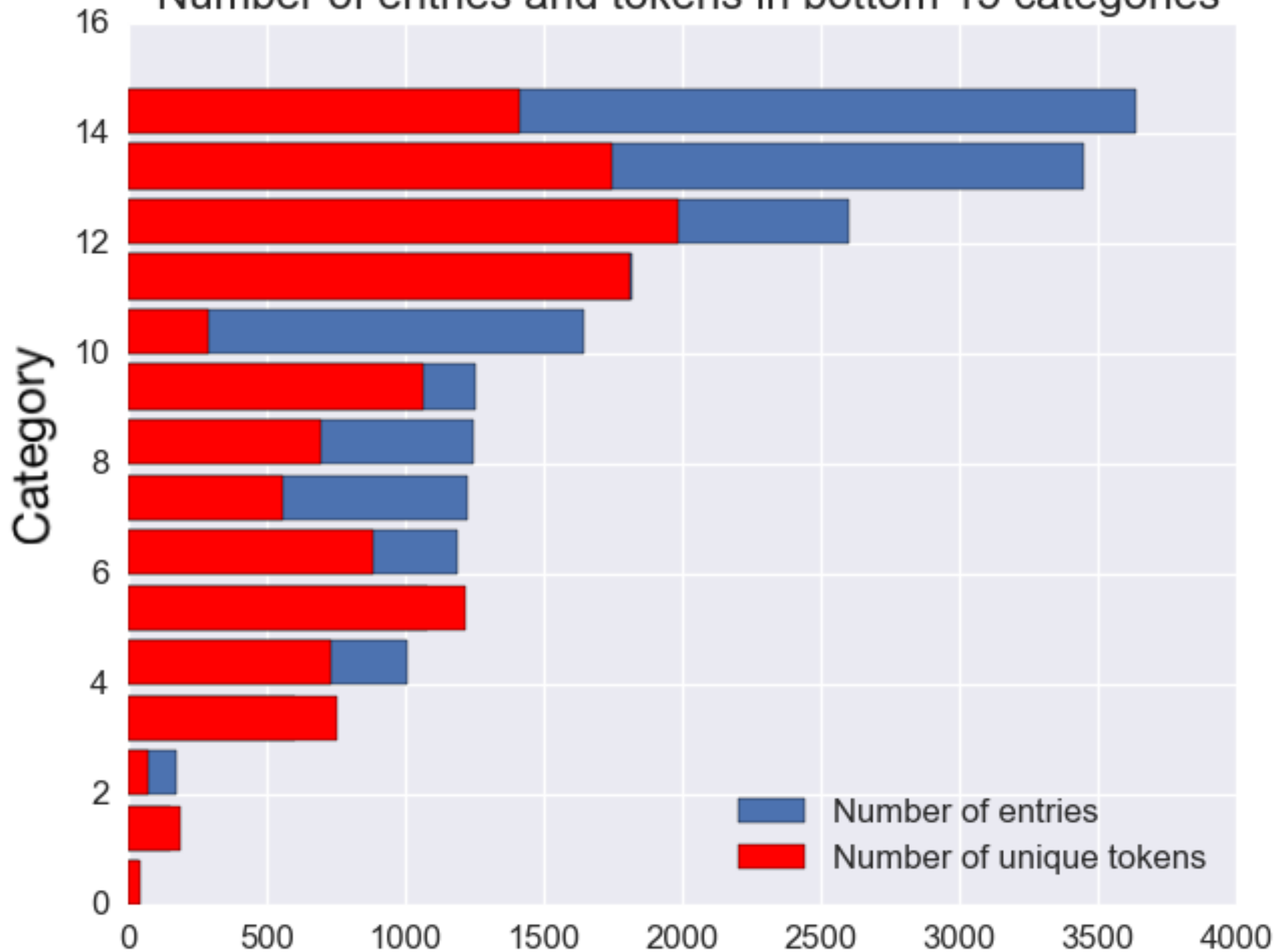


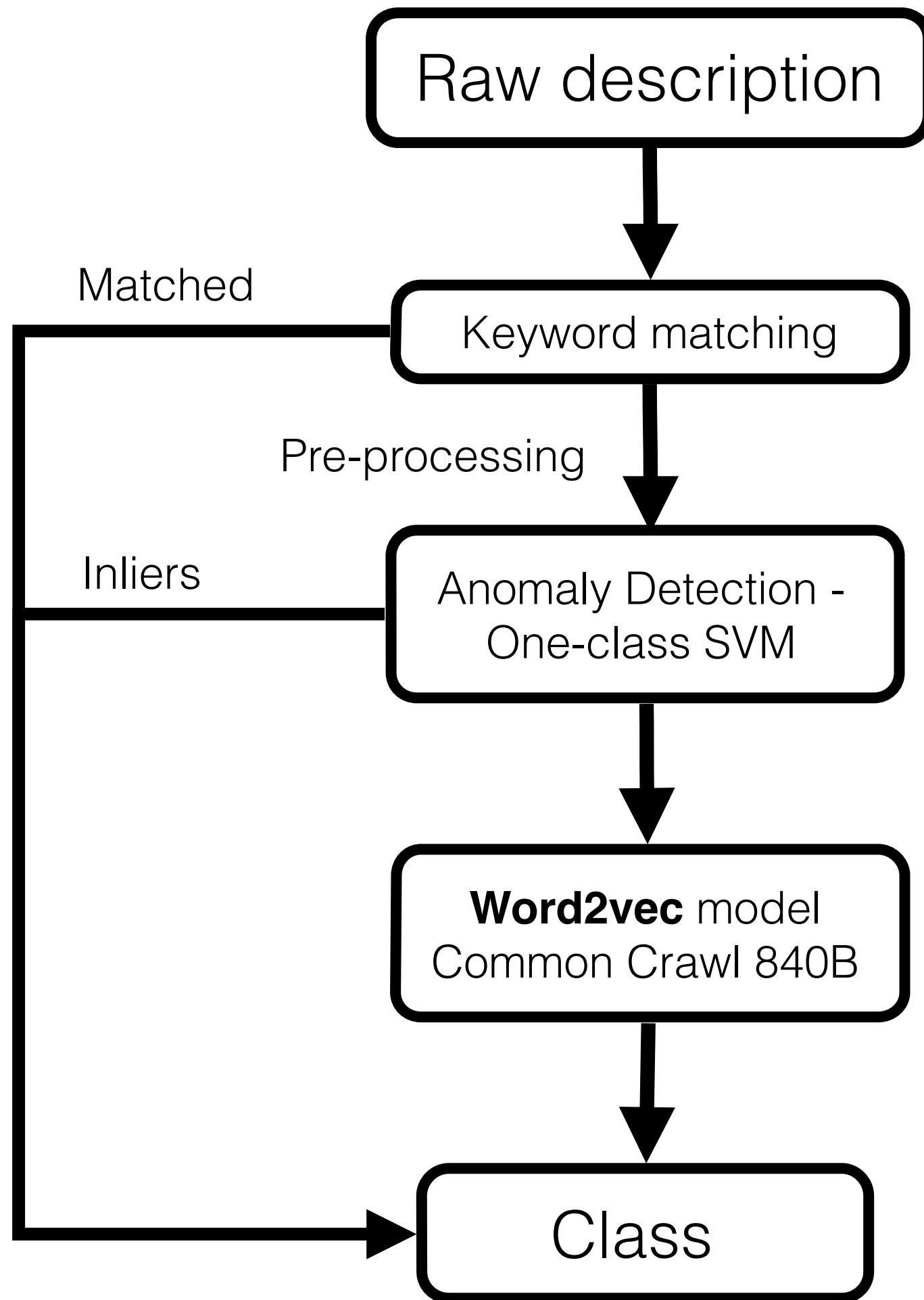
HVAC

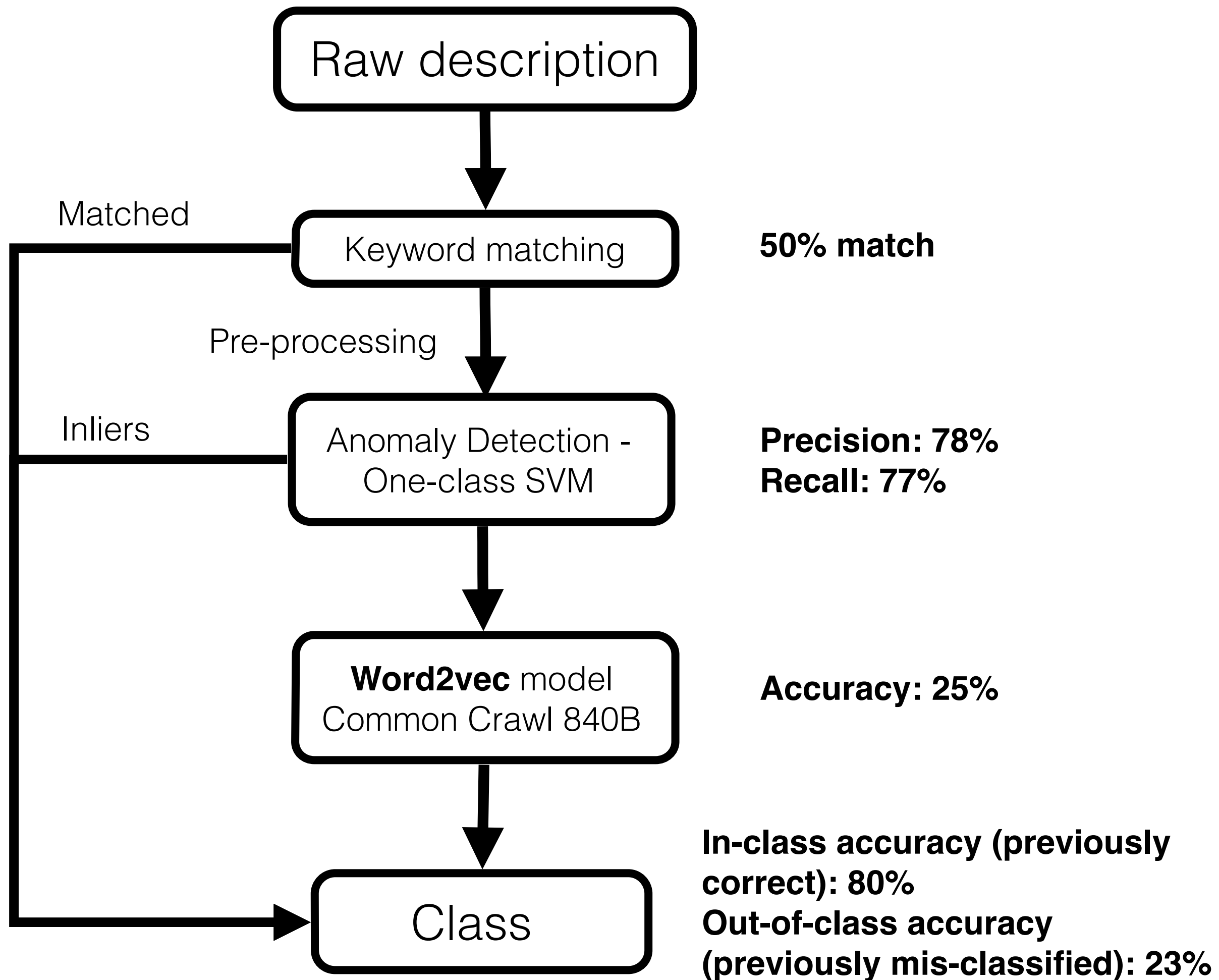
Number of entries and tokens in top 15 categories



Number of entries and tokens in bottom 15 categories







Recommendations to client

- Merge classes with semantically similar descriptions
e.g. “HVAC” and “Heating”
- Use more specific class descriptions
e.g. “Roof Repair” instead of “Roof”

Ivan Teo

Computational biophysics

U of Illinois at Urbana-Champaign

<https://ivanteokh.github.io/posts/insight>

