# Establishing a Strong Baseline for Privacy Policy Classification

* Multi-label classification problem with 12 classes based on the OPP-115 dataset.
* Two gold standards from OPP-115:
  + Majority-vote-based: two or more exports agree on a label.
  + Union-based: contains all export annotations.
* Reason of dividing into 12 categories:
  + The annotations in OPP-115 dataset are in two levels:
    - 10 high level categories.
    - 22 distinct attributes.

The paper solves the multi-label classification problem with 12 classes. It has two gold standards from OPP-115. The OPP-115 is a dataset that contains 115 privacy policies at paragraph level, each of which includes fine-grained annotations from 3 exports.

The two gold standards include majority-vote-based that means two or more exports agree on a label of one paragraph, and union-based that one paragraph contains all export annotations.

Why the paper divided it into 12 categories? The annotations in OPP-115 dataset are in two levels: 10 high level categories and 22 distinct attributes. The paper remained 10 high level categories and divided the “Others” category into other three categories later.