

# AlpacaTag: An Active Learning-based Crowd Annotation Framework for Sequence Tagging

Bill Yuchen Lin<sup>\*1</sup>, Dong-Ho Lee<sup>\*1</sup>, Frank F. Xu<sup>2</sup>, Ouyu Lan<sup>1</sup>, and Xiang Ren<sup>1</sup>

<sup>1</sup> University of Southern California <sup>2</sup> Carnegie Mellon University

<http://inklab.usc.edu/AlpacaTag>

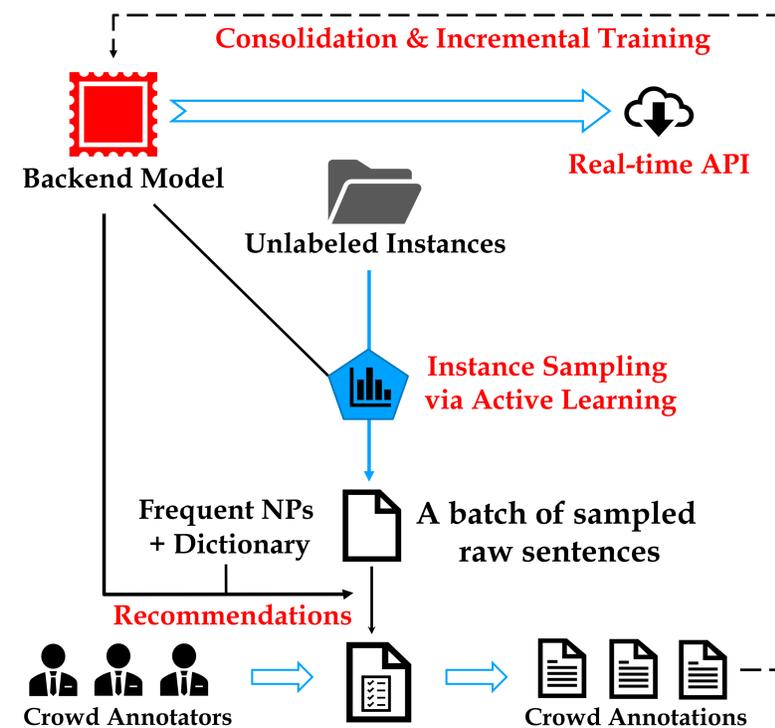


## Easy-to-use Intelligent Recommendations

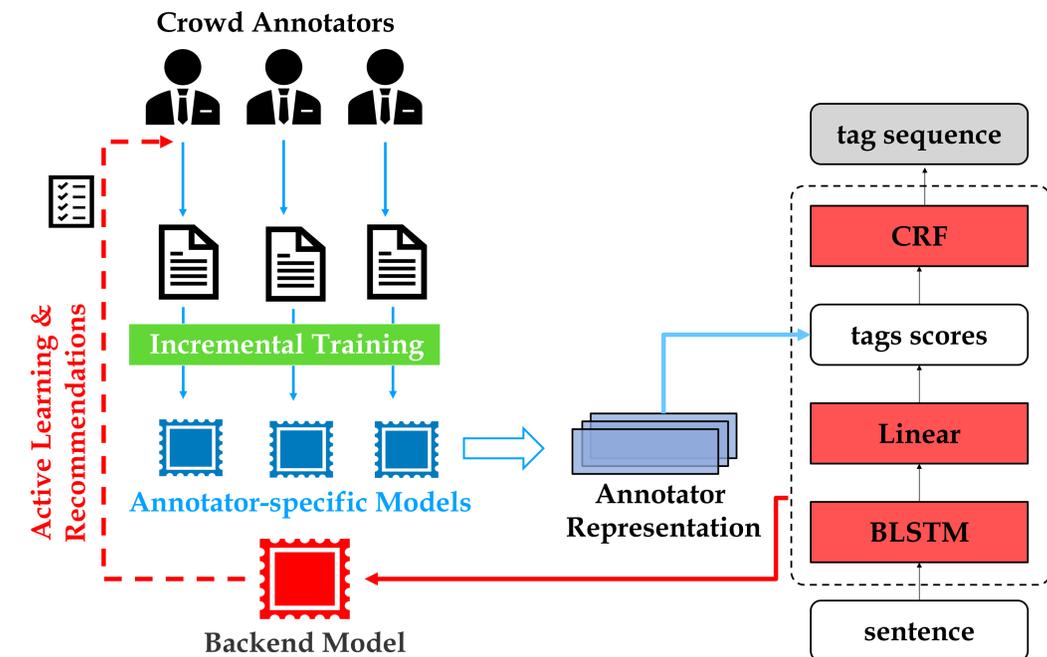
dynamically adjusting the order of samples to annotate

inference confidence of backend models

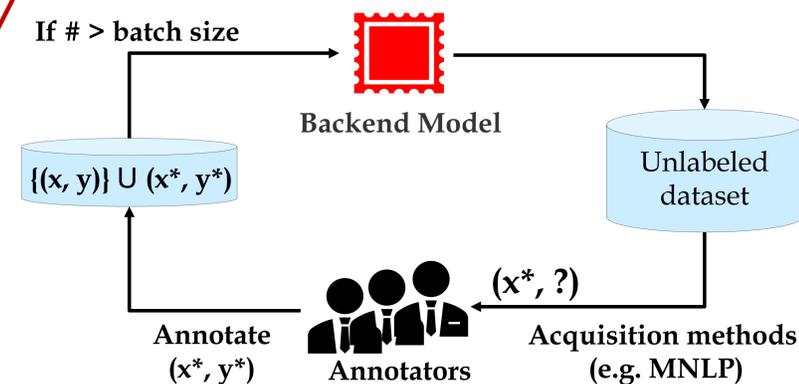
## Overall Workflow of AlpacaTag



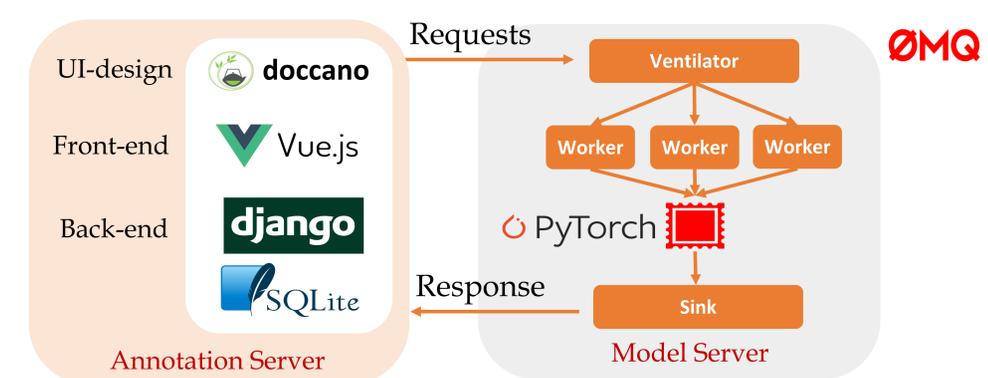
## Learning Backend Model with Consolidation



## Active Learning for Instance Sampling



## Framework Implementations



## Key Features

- Active Intelligent Recommendation**  
Dynamically suggesting annotations and actively sampling the most informative unlabeled instances.
- Annotation Consolidation**  
Enhancing inter-annotator agreement by consolidating multiple personal backend models.
- Real-time model deployment**  
Users can deploy backend models in downstream system via APIs while annotators are tagging data.

## Extensible Configurations

- Embedding**  Glove  Word2Vec  Fasttext  ELMO  BERT
- Recommendation**  Noun Chunk  Online Learning  Dictionary Match
- Active Learning**  Random  MNL  MNL + MC  MNL + BB

- Users can optionally enable each source in the Settings.
- Final recommendations are merged from three options with the priority:

“Noun Chunk < Model Inference < Dictionary Match”



Scan me