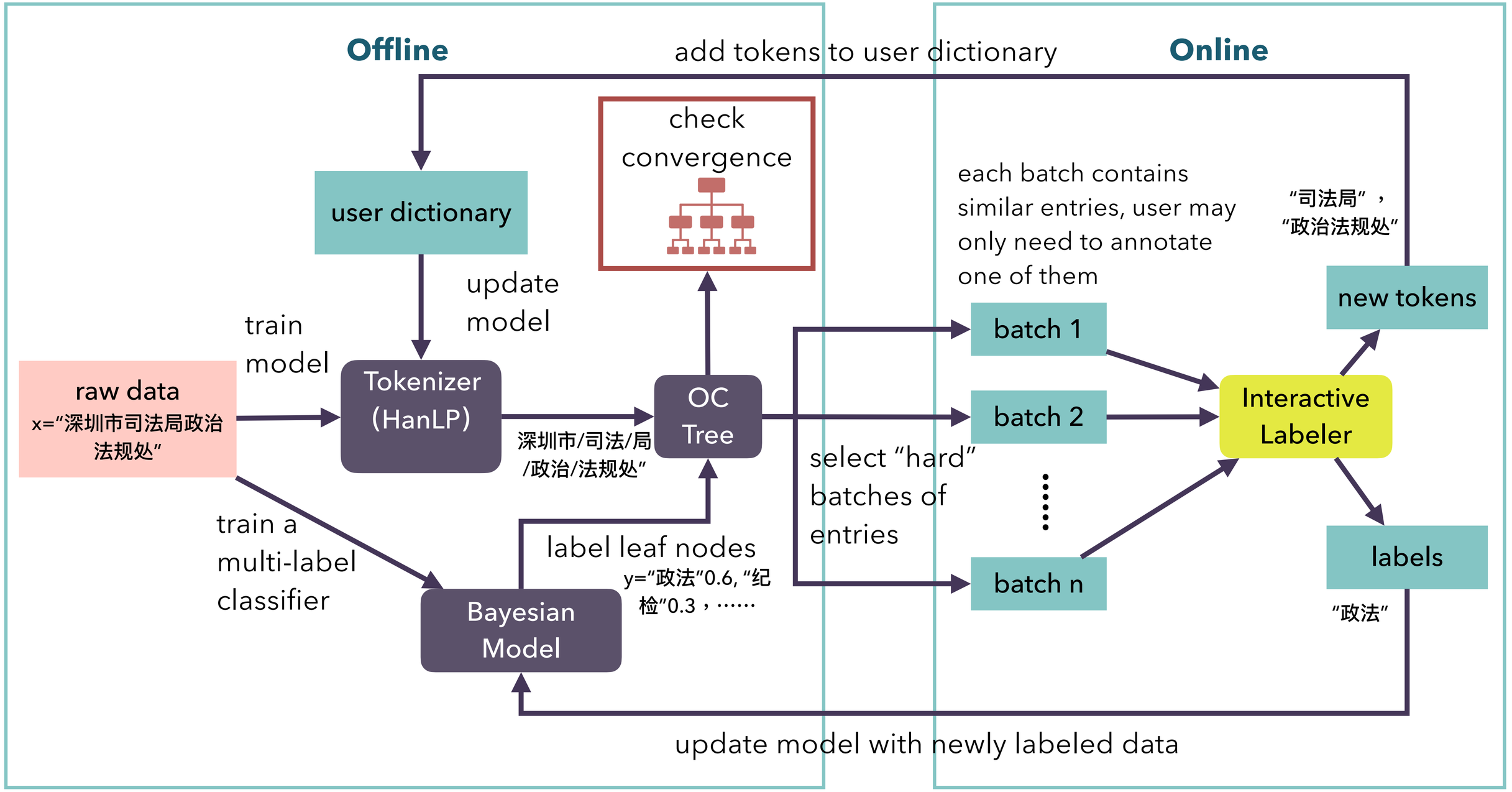
An Active Learning System for Resume Data Tokenization & Labeling



Short-Term Goals

- OC tree visualization
- career profiling
  - define a similarity metric based on multiple criteria (e.g. career history, social connection etc)
  - retrieval/clustering based on similarity metric
  - **attribute-based** (learn all “tags” associated with a person’s past experience) or **trajectory-based** (focus on transitions between careers, use time warpping to match career trajectories from different time periods)
- career path prediction

短期目标分工（暂定）

- 家硕：tree visualization, multi-label classification
- Calvin：HanLP, online model update
- 志威：interactive annotation platform, batch selection (based on HanLP tokenizer’s confidence score or heuristics)

Long Term Applications

- fuzzy search using natural language eg. “具有部队与基层经历和工程背景的人员“
- recommendation for a group of positions with specific team composition and group dynamic
- social computing topics: e.g. review career networks and their social dynamics of different cities, identify key events in people’s career path