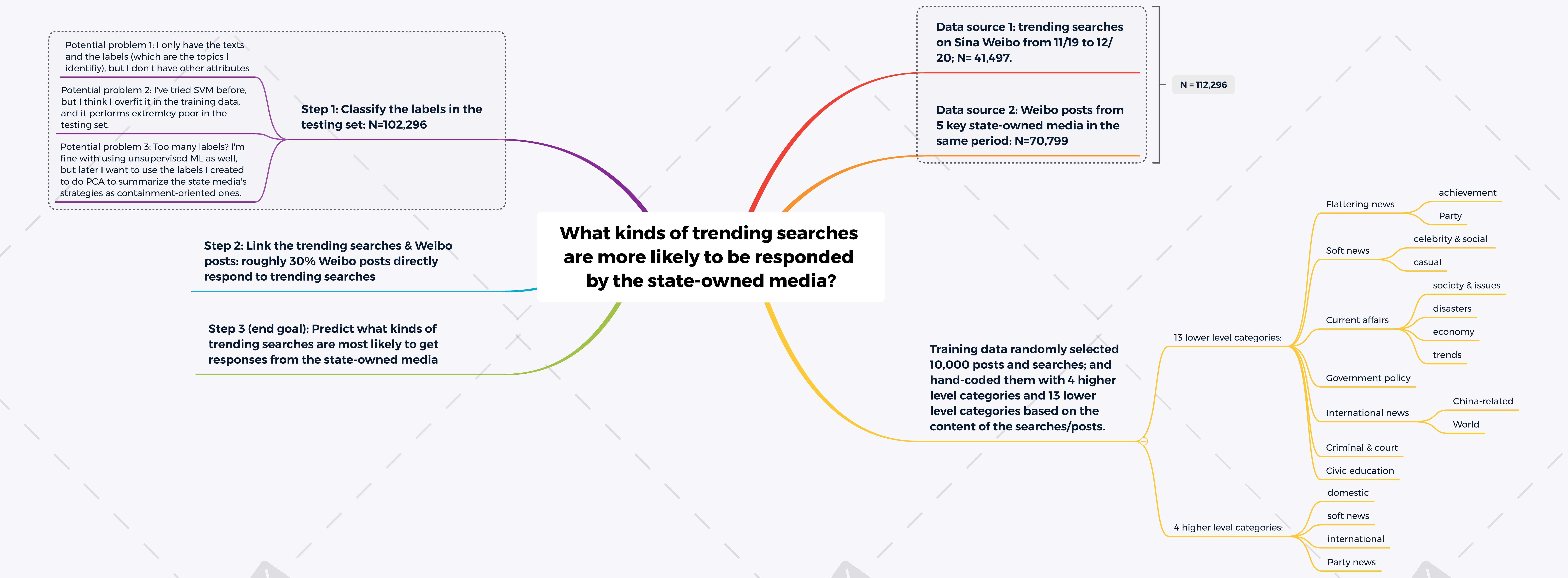

What kinds of trending searches are more likely to be responded by the state-owned media?

- **Data source 1: trending searches on Sina Weibo from 11/19 to 12/20; N= 41,497.**
- **Data source 2: Weibo posts from 5 key state-owned media in the same period: N=70,799**
  - N = 112,296
- **Training data randomly selected 10,000 posts and searches; and hand-coded them with 4 higher level categories and 13 lower level categories based on the content of the searches/posts.**
  - 13 lower level categories:
    - Flattering news
      - achievement
      - Party
    - Soft news
      - celebrity & social
      - casual
    - Current affairs
      - society & issues
      - disasters
      - economy
      - trends
    - Government policy
    - International news
      - China-related
      - World
    - Criminal & court
    - Civic education
  - 4 higher level categories:
    - domestic
    - soft news
    - international
    - Party news
- **Step 1: Classify the labels in the testing set: N=102,296**
  - Potential problem 1: I only have the texts and the labels (which are the topics I identifiy), but I don't have other attributes
  - Potential problem 2: I've tried SVM before, but I think I overfit it in the training data, and it performs extremley poor in the testing set.
  - Potential problem 3: Too many labels? I'm fine with using unsupervised ML as well, but later I want to use the labels I created to do PCA to summarize the state media's strategies as containment-oriented ones.
- **Step 2: Link the trending searches & Weibo posts: roughly 30% Weibo posts directly respond to trending searches**
- **Step 3 (end goal): Predict what kinds of trending searches are most likely to get responses from the state-owned media**