Advanced Internet Computing

WS 2013/14

Topic 2: Crowdsourcing

Bernhard Gößwein
David Riepl
Robert Kapeller

Topic description

- Plan and implement Sentiment analysis for products and companies mentioned in article-feeds
  - ◆ Gather feed-contents
  - ◆ Crowd: Identify keywords for products and/or companies
  - ◆ Crowd: Rate keywords in conjunction with article
  - ◆ REST interface for sentiment querying
- Implement Crowd sourcing simulator
  - ◆ (MobileWorks was not reliable...)

Technology

- Python
- Django
  - ◆ REST / JSON / Tastypie / ...
- Feedparser
- SQLLite
- Bootstrap (for GUI, mockups)

Our solution

- 2 Tasks
  - ◆ Identify keywords
    - Feed-fetch automated via cron jobs (yahoo finance)
  - ◆ Rate keywords
    - Customer driven (Order)
    - Task-count depend on customer budget and article length
- Sentiment analysis (graphically)
  - ◆ Average sentiment per keyword
  - ◆ Detailed view
  - ◆ Graphical analysis
- GC
  - ◆ Delete old unfinished tasks

Quality control
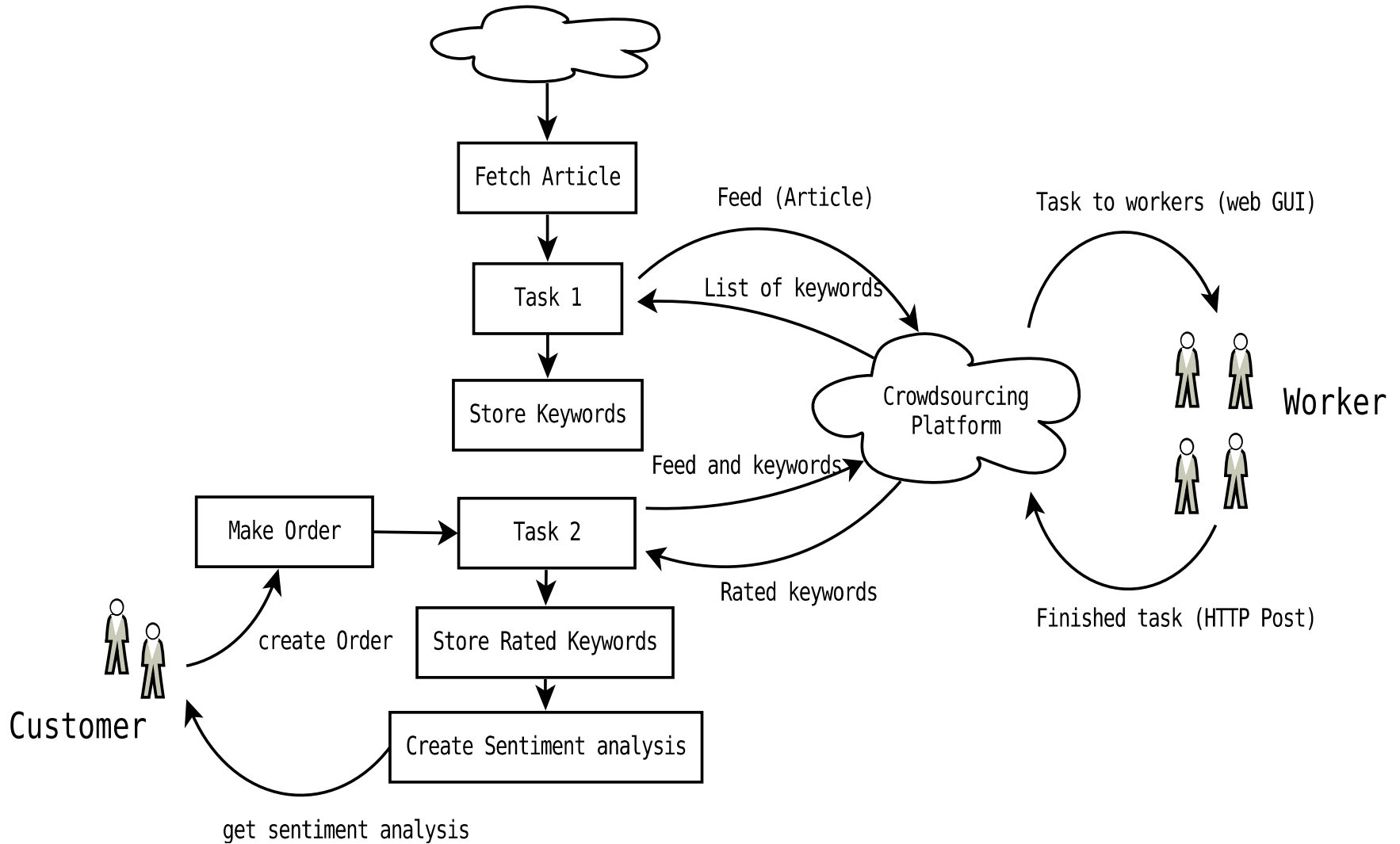
- Task 1 – keyword detection
  - Worker rated with negatively if:
    - keyword not in original text (currently no word - normalization). Keyword is not accepted.
    - Majority voting on keyword/category combination
- Task 2- sentiment voting
  - Spam-prevention: if distance to median >= 3 worker rated negatively
- Redundancy
  - Task 1: Each feed posted 3 times
  - Task 2: Depends on budget

DB Model
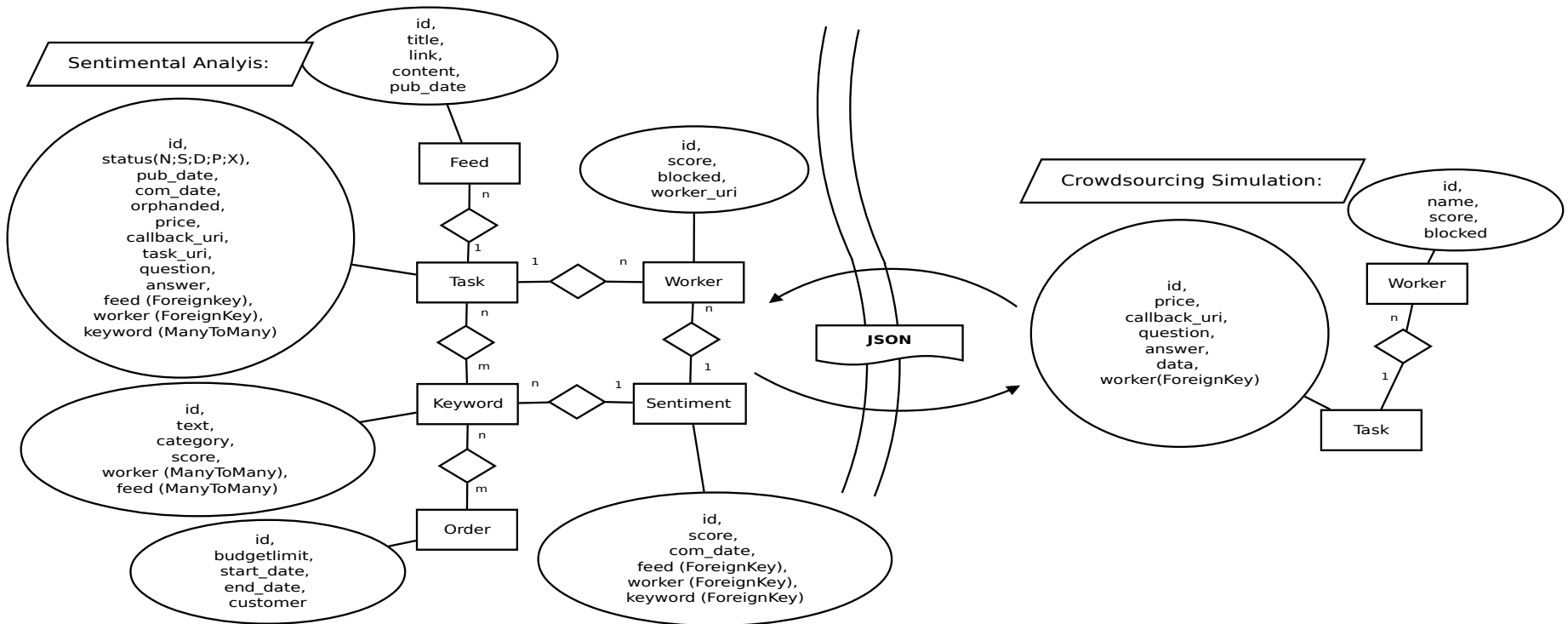
Thank you!

Now lets see it in action!