

A Clustering and Recommendation System for Opinion Articles in Turkish newspapers

Üstün Özgür

Aim

- Detect common themes in articles
- Discover themes of current day/week/month
- Cluster today's opinion papers
- Recommend related opinion articles/columnists

Motivation/Importance

- Zeitgeist: pulse of the moment
- See the big picture: Discover themes of the day
- Find supporting opinions
- Find opposing opinions
- Discover new authors
- Recap last week/month instead of day

Methodology

- Scraping and content extraction
- Indexing and stemming (ES+custom stemmer)
- Compare at least two clustering techniques
 - One hierarchical and one flat
- Recommendation system
 - content-based approach (probably)
 - Ask the user a few questions initially
 - Allow user to like/dislike author or article
 - collaborative (not likely, not enough users)

Expected Results

- Current articles should be classified in K clusters, where $K < 5$ or 10
- Clustering should be all automatic, no manual intervention
- Articles should be recommended to the user after reading n articles. Recommendation should adapt user.
- Recommended articles should be current, but could be a week/month old