This is the program flow chart

QUERIES

NEWS

Broadcast

For the efficiency, NEWS will filtering with the query
terms. I will handle the contents only has query terms.
that is the articles containing the query terms

Process and Filtering

Articles Containing the Query Terms

Get TFs for Query Terms per docs

Query term TFs (qterm x doc)

Group by term

Calculate TFs

Calculate DPH

Query term DPH (qterm x doc)

Accumulators

Corpus Stats

Group by Query

Calculate Query DPH

Filter to top 50

Get Articles in Rankings

Top 50 rankings (Query x 50)

Matched Article (Query x 50)

Redundancy check

Top 10 Result (Query x 10)