

start

Phase 1

Candidate label
discovery

Phase 2

Cluster
discovery

Potential good
cluster labels

Model of dominant
topics

Phase 3

Pattern phrase
selection

Pattern phrases

Document assignment
to pattern phrases

Pruning and
postprocessing

end

