

Data collection



Data cleaning and  
pre-processing



Document  
embeddings



Self-organizing map  
(SOM)



Document  
exploration interface