

Potrebno je osmisliti i implementirati algoritam za klasterizaciju dokumenata.

Porediti klasične – uobičajene mjere rastojanja sa Mahalanobis rastojanjem.

Uporediti načine za generisanje reprezentacije klastera.

Ispitati efekat slučajne modifikacije dokumenta na konačne performanse (pogledati CURE algoritam).

Ispitati da li dokumenti koji ostanu van svih klastera predstavljaju izuzetke ili novi klaster/klasu.