

Topic Mining Revisited

INPUT: C, k, V

OUTPUT: $\{ \theta_1, \dots, \theta_k \}, \{ \pi_{i1}, \dots, \pi_{ik} \}$

Text Data



θ_1

sports 0.02
game 0.01
basketball 0.005
football 0.004
...

θ_2

travel 0.05
attraction 0.03
trip 0.01
...

...

θ_k

science 0.04
scientist 0.03
spaceship 0.006
...

Doc 1

30%

π_{11}

Doc 2

$\pi_{21}=0\%$

...

Doc N

$\pi_{N1}=0\%$

12%

π_{12}

π_{22}

π_{N2}

8%

π_{1k}

π_{2k}

π_{Nk}