Latent Search Indexing

Mathematical Modelling

Nik Jenič, Tian Ključanin, Maša Uhan

Problem Introduction

- Finding relevant documents according to our search

Solution

- LSI - Latent Semantic Indexing

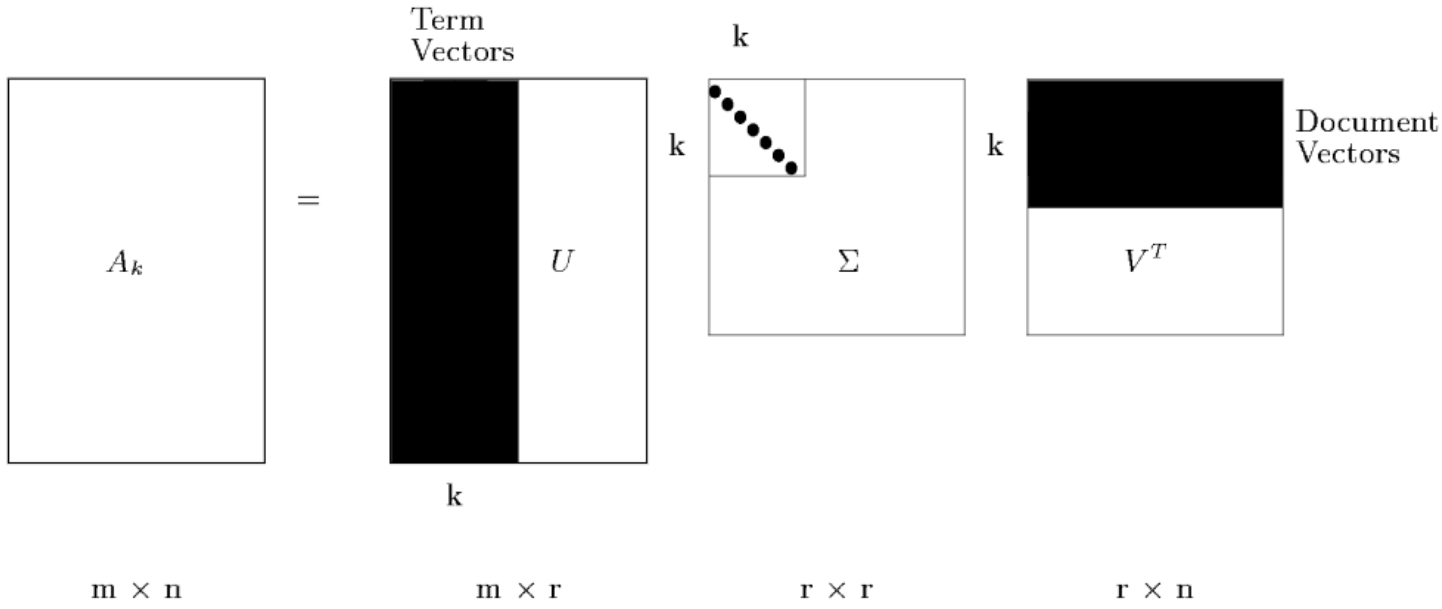

Figure: Mathematical representation of A_k

Optimization

- Giving words different weights
- Different ways of calculating the weights

$$a_{ij} = L_{ij} \cdot G_i$$

$$L_{ij} = log(1 + f_{ij}), \quad G_i = 1 - \sum_j \frac{p_{ij} log(p_{ij})}{logn}, \quad p_{ij} = \frac{f_{ij}}{g_{f_i}}$$

Additional Improvements to the Solution

- Adding new documents without recalculation of SVD
- Adding new words without recalculation of SVD

Results

- NE PUSTIT PRAZNO
-

Discussion

- NE PUSTIT PRAZNO

References

- Source for Figure 1: M. W. Berry, S.T. Dumais, G.W. O'Brien, Michael W. Berry, Susan T. Dumais, and Gavin. Using linear algebra for intelligent information retrieval. SIAM Review, 37:573–595, 1995