Singular Value Decomposition

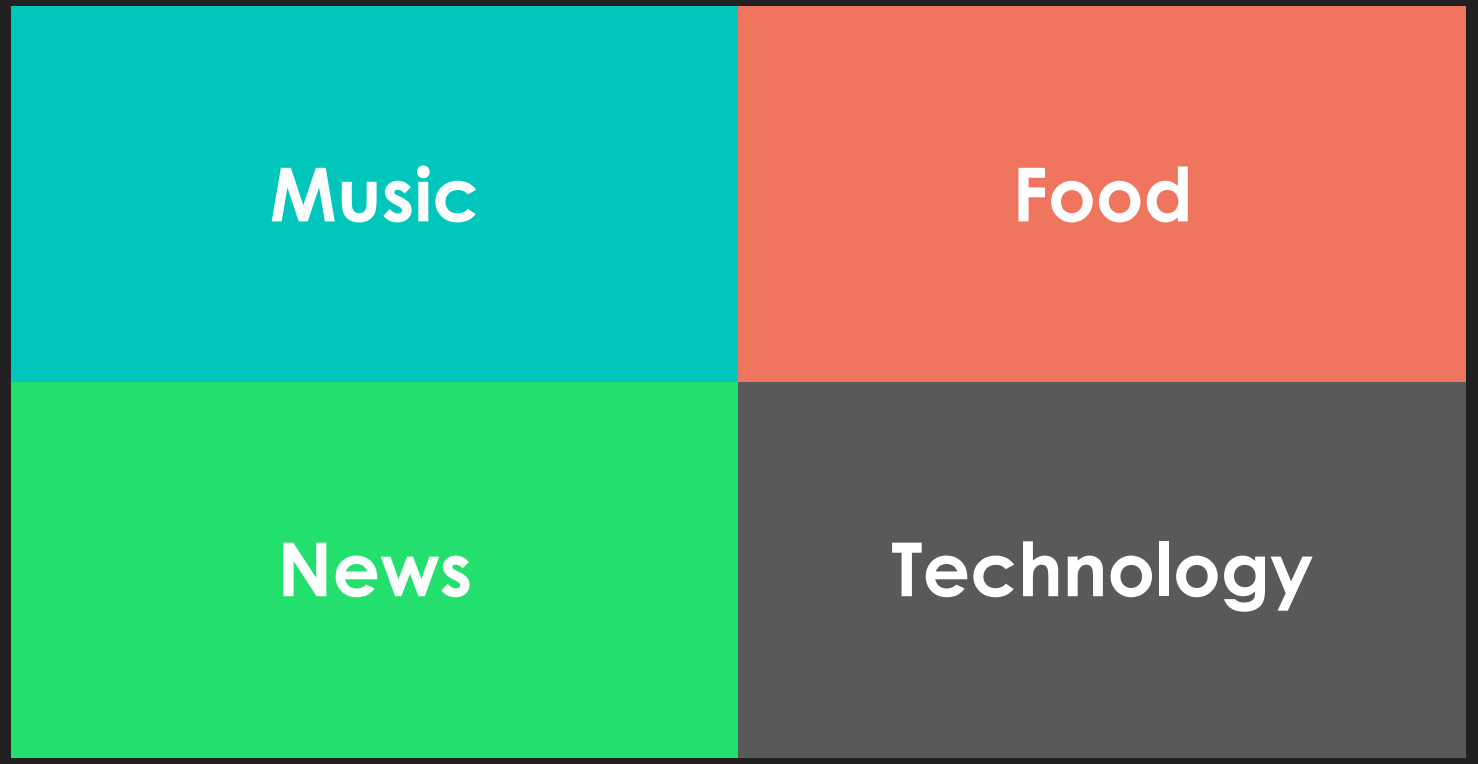
Introduction to SVD and LSA

Latent Semantic Analysis

"Latent Semantic Analysis is a technique of analysing relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms."

- Wikipedia

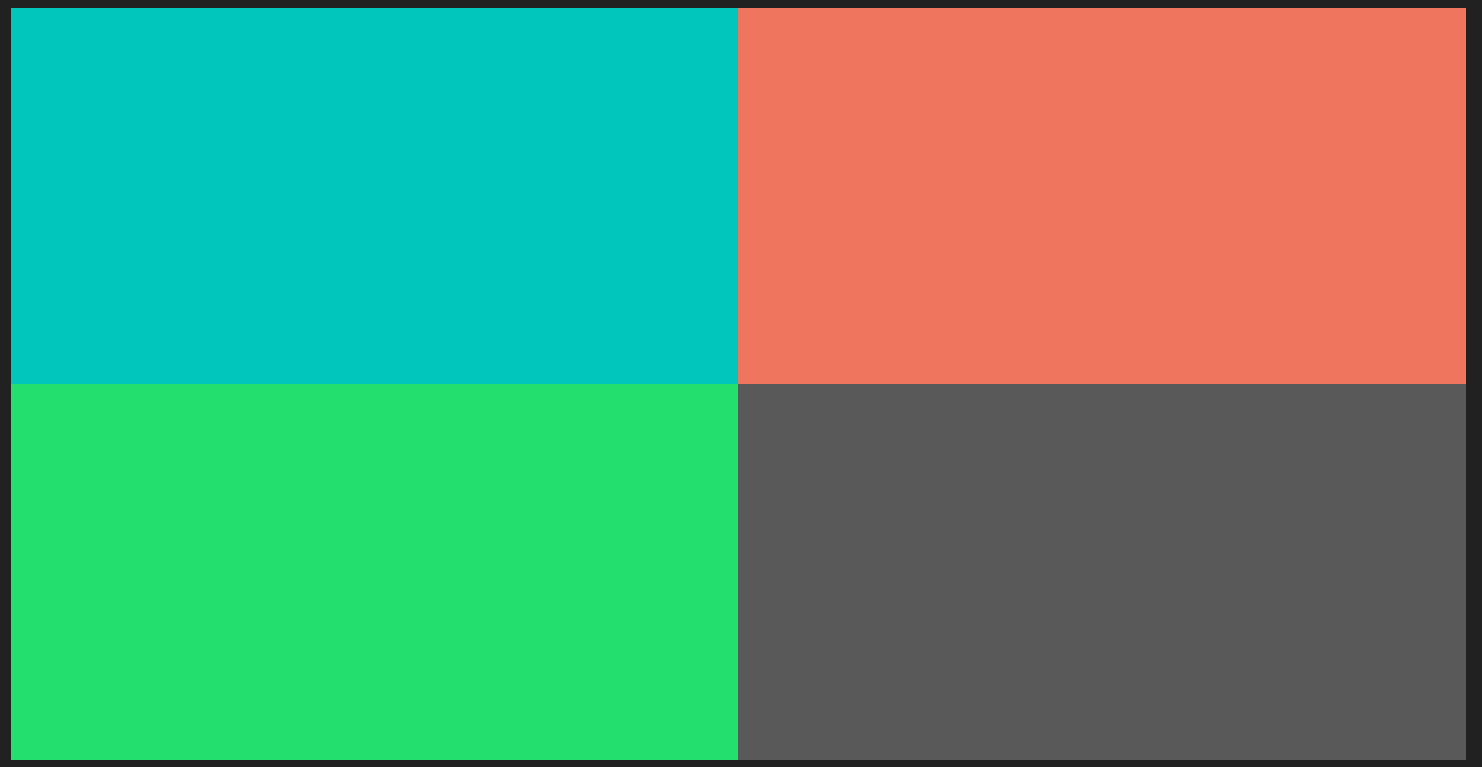
Latent Semantic Analysis



Latent Semantic Analysis

1. Article 1
2. Article 2
3. Article 3
4. Article 4
5. Article 5
6. Article 6

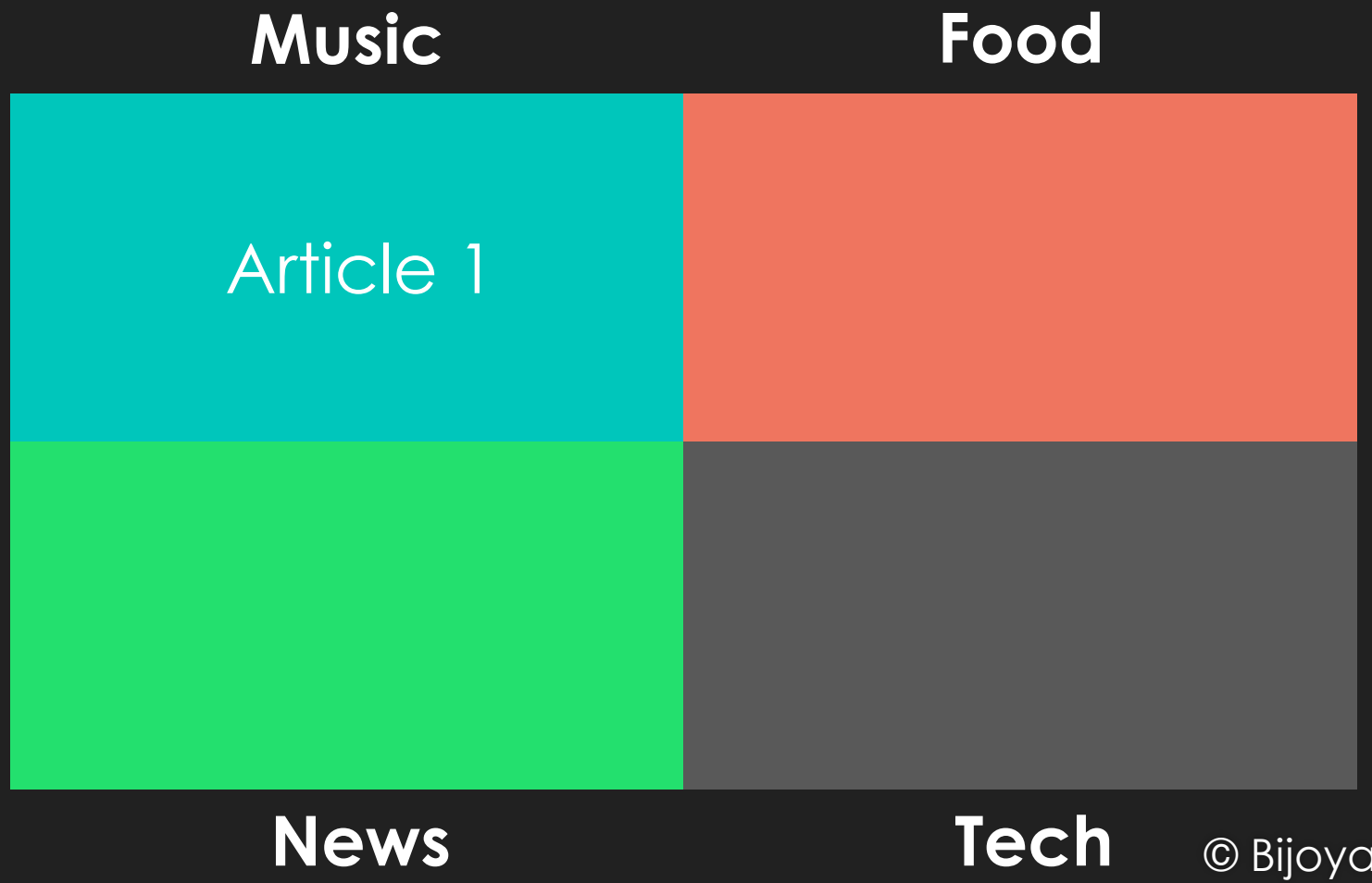
……



Latent Semantic Analysis

1.
2. Article 2
3. Article 3
4. Article 4
5. Article 5
6. Article 6

……



Latent Semantic Analysis

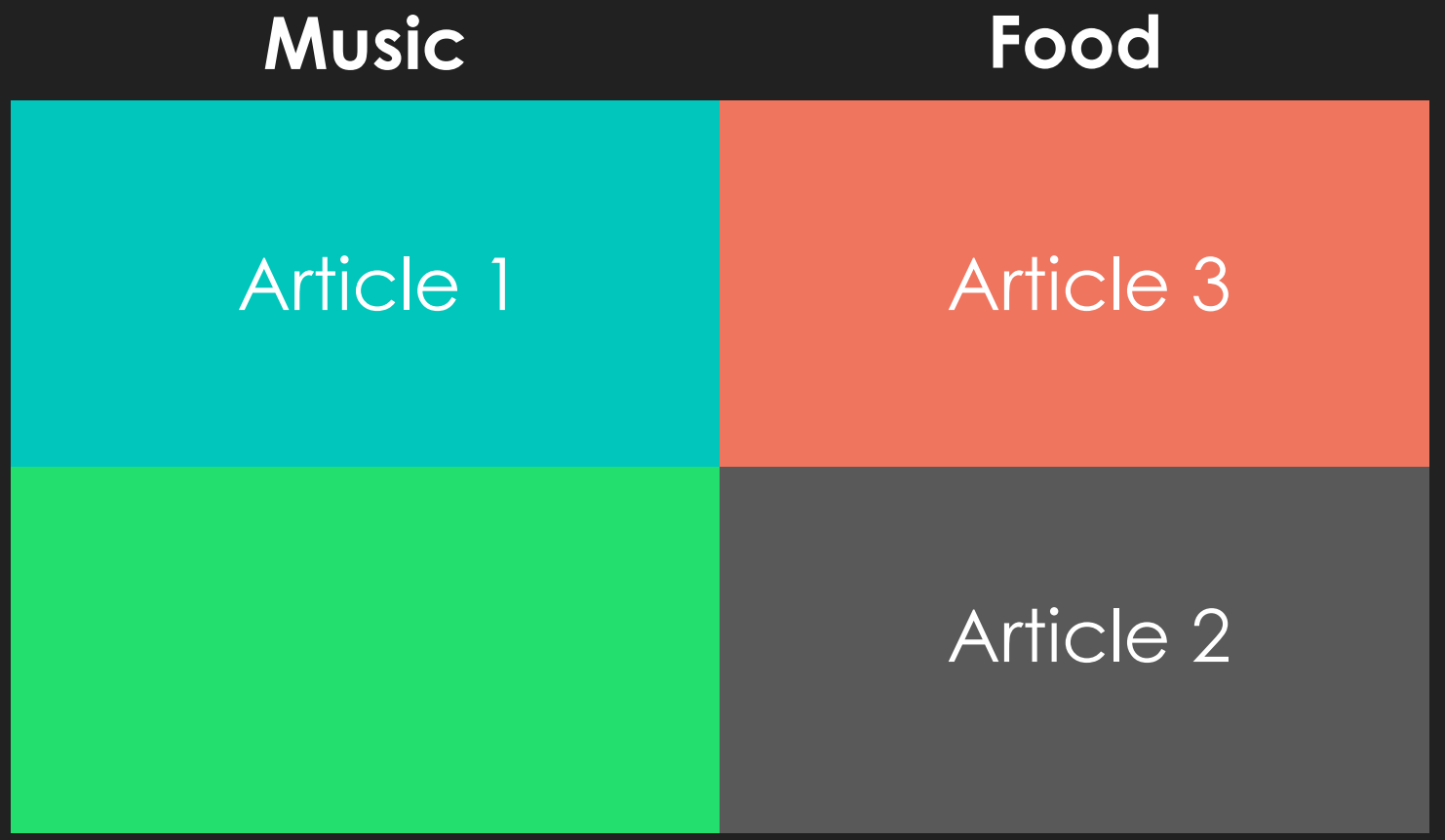
1.
2.
3. Article 3
4. Article 4
5. Article 5
6. Article 6

……

Latent Semantic Analysis

1.
2.
3.
4. Article 4
5. Article 5
6. Article 6

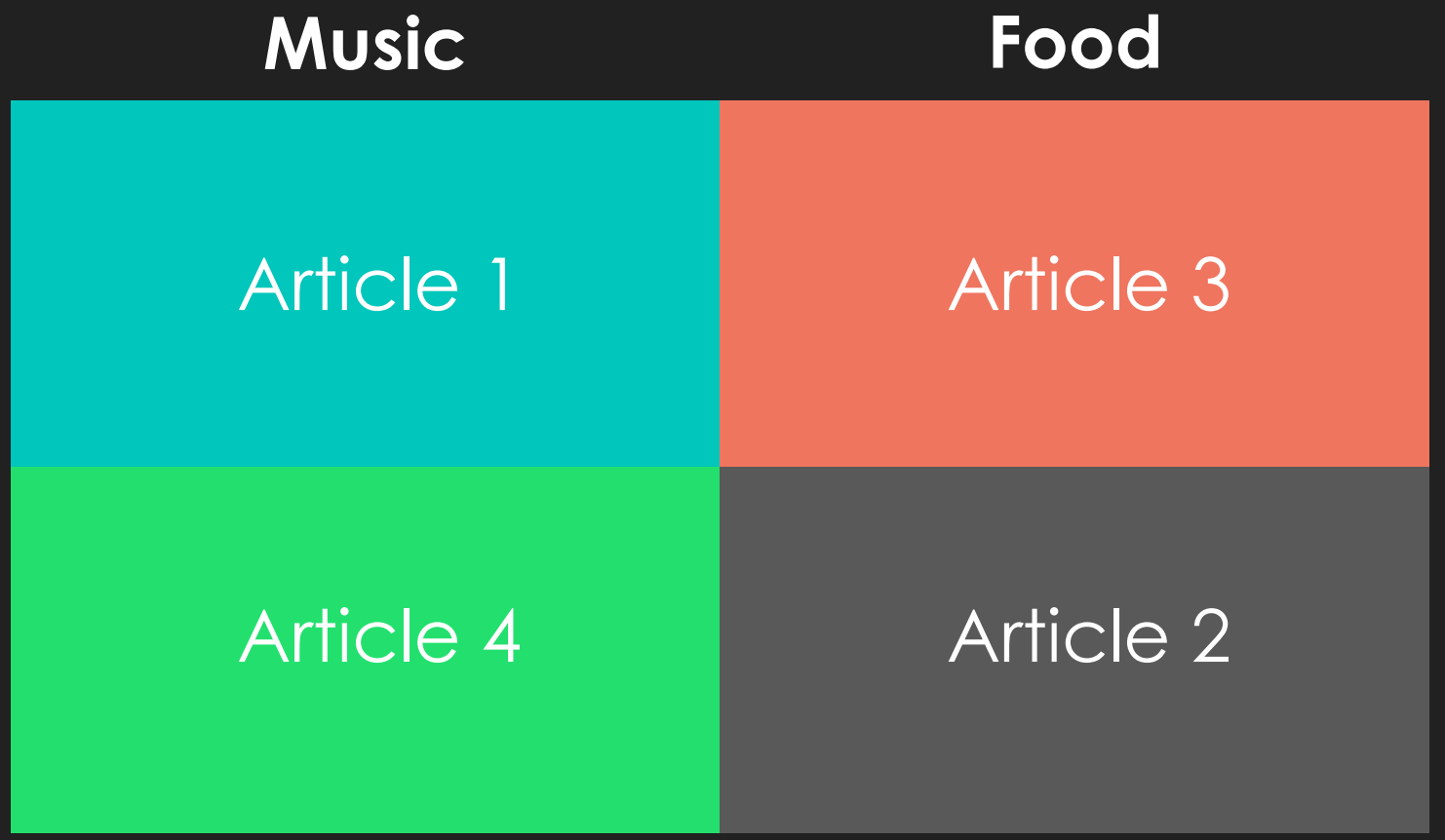
……



Latent Semantic Analysis

1.
2.
3.
4.
5. Article 5
6. Article 6

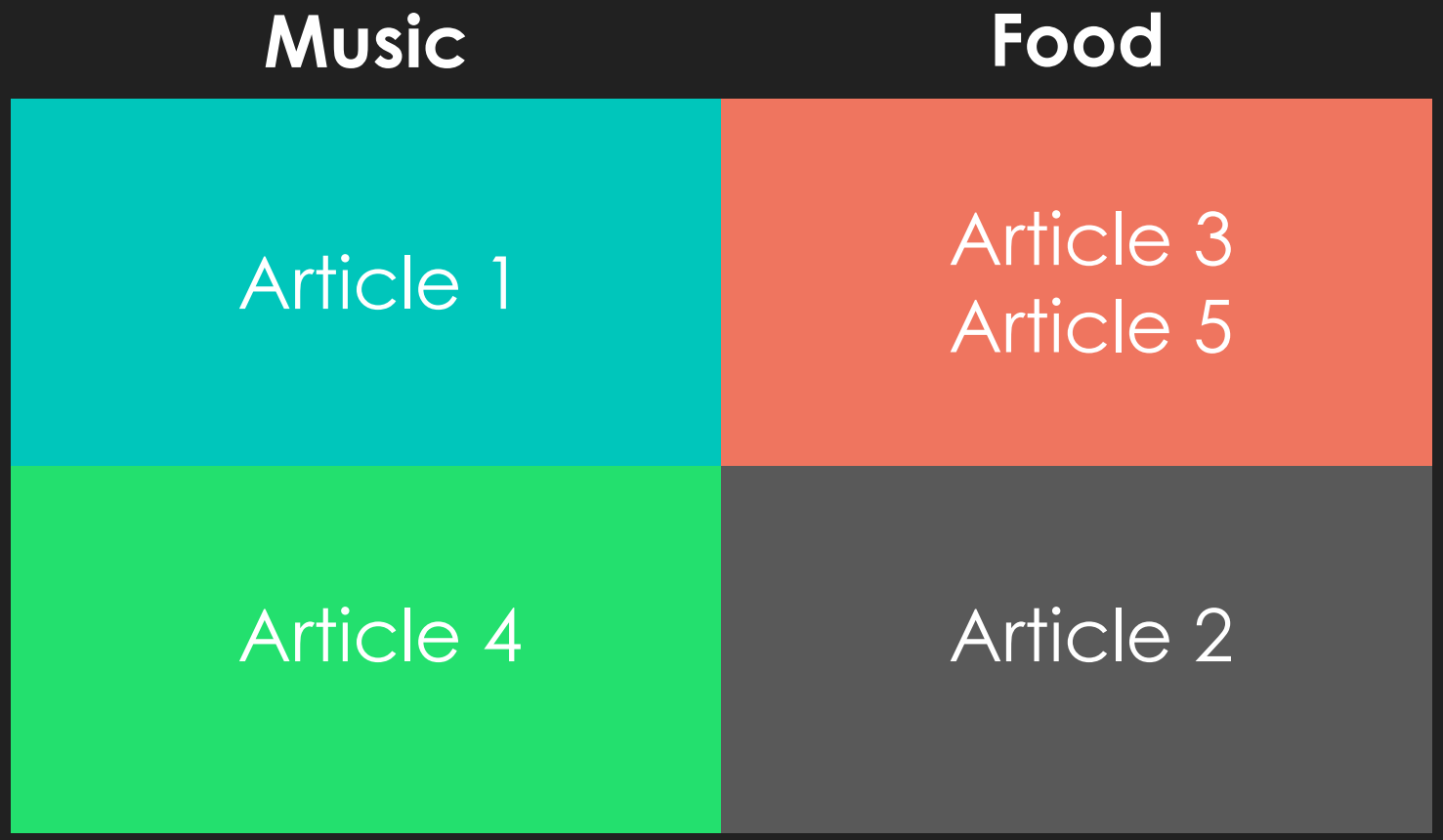
……



Latent Semantic Analysis

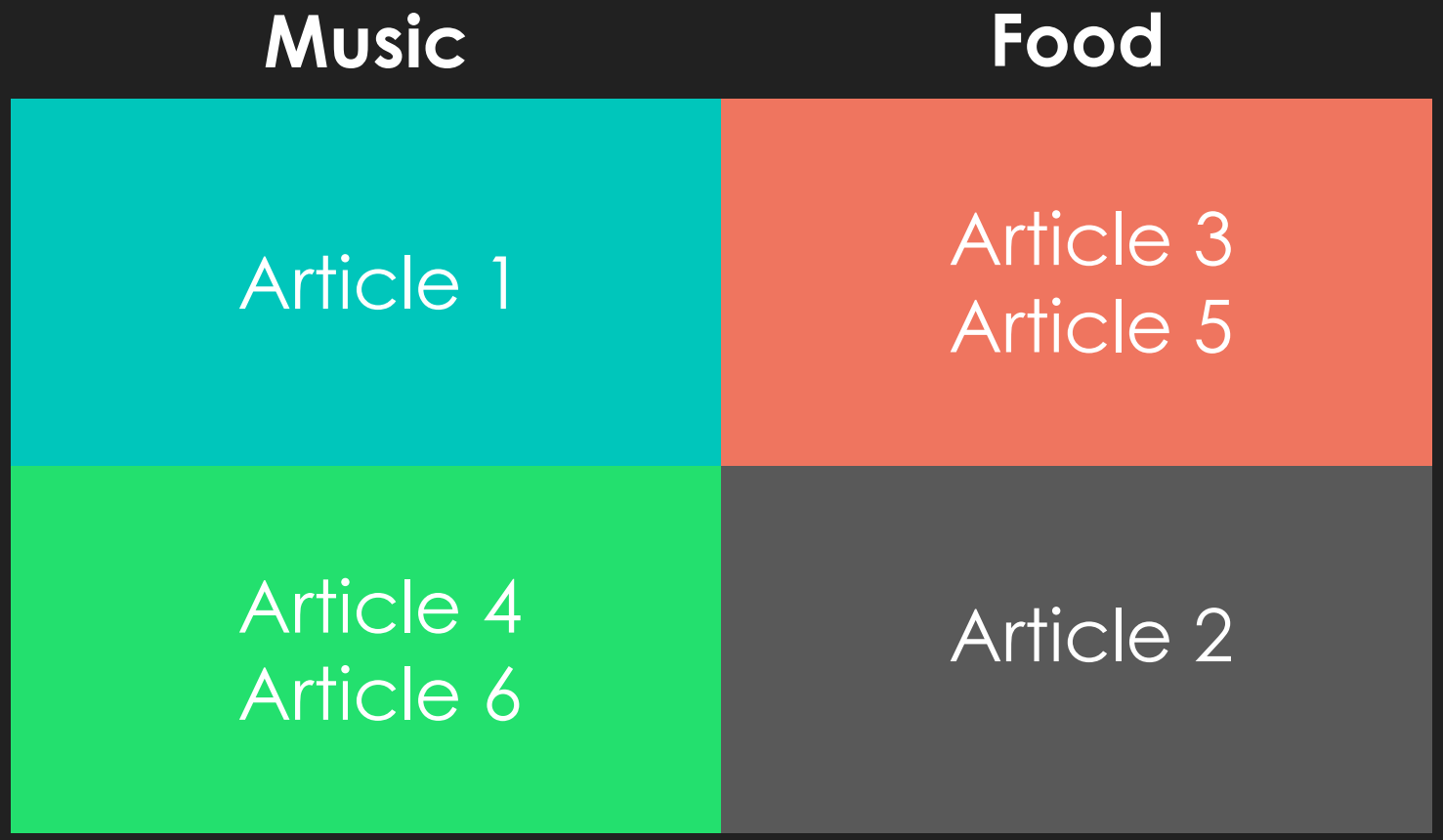
1.
2.
3.
4.
5.
6. Article 6

……



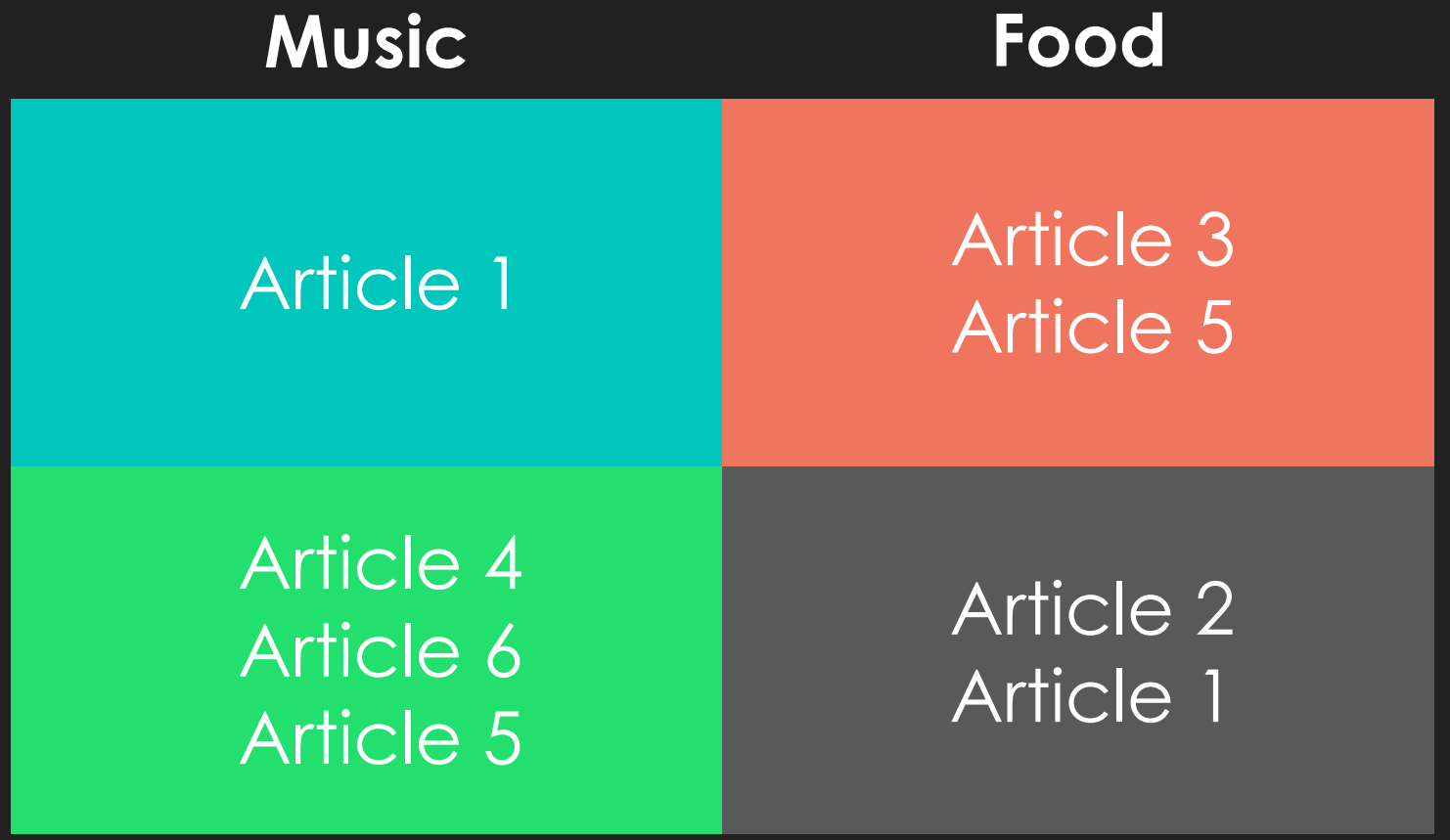
Latent Semantic Analysis

1.
2.
3.
4.
5.
6.
……



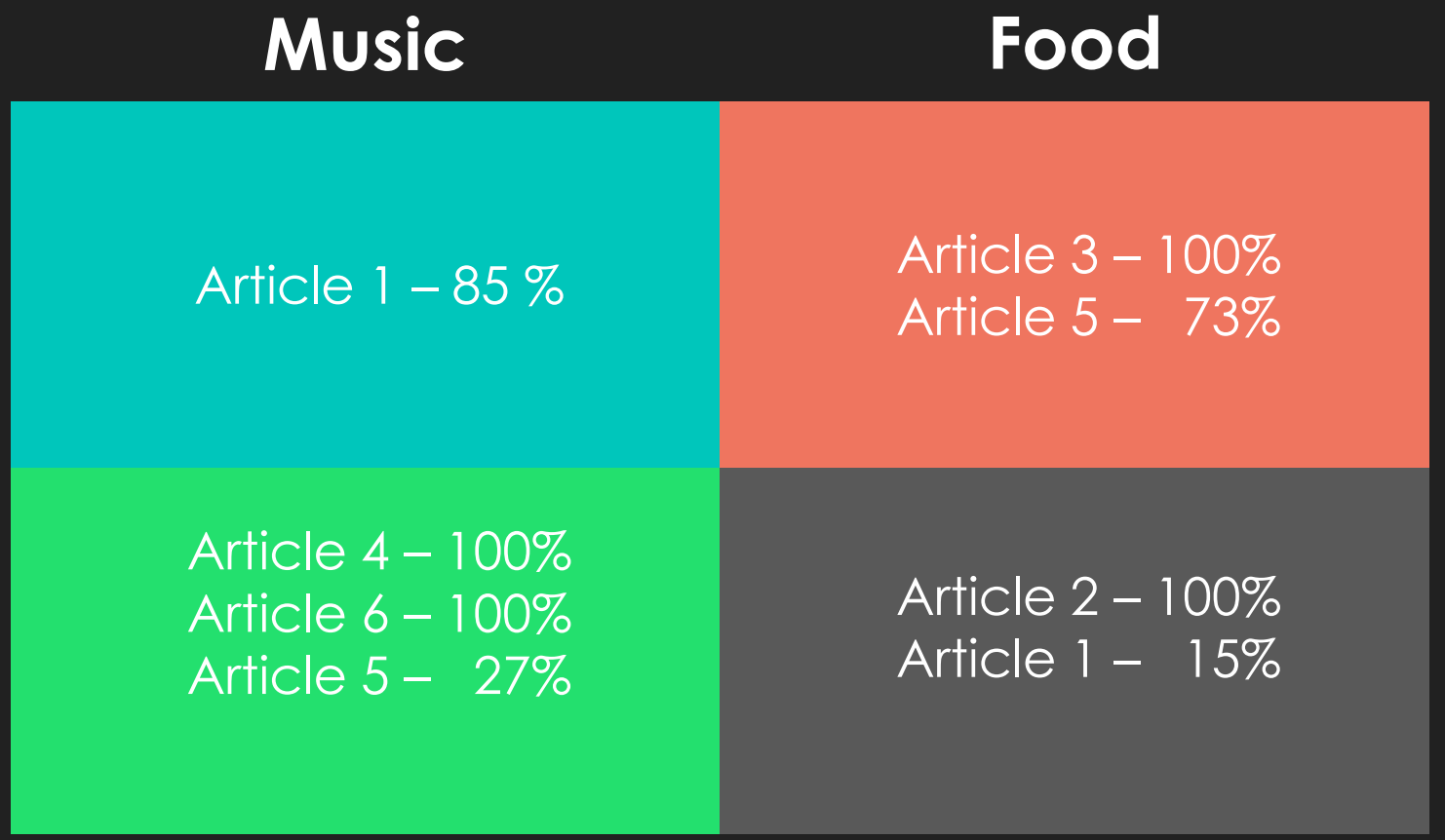
Latent Semantic Analysis

1.
2.
3.
4.
5.
6.
……



Latent Semantic Analysis

1.
2.
3.
4.
5.
6.
……



Bag Of Words Model

Words/Documents	going	to	today	i	am	it	is	rain	not	outside
1	1	1	1	0	0	1	1	1	0	0
2	1	0	1	1	1	0	0	0	1	1
3	1	1	0	1	1	0	0	0	0	0

M x N matrix

M = Number of Rows/Documents

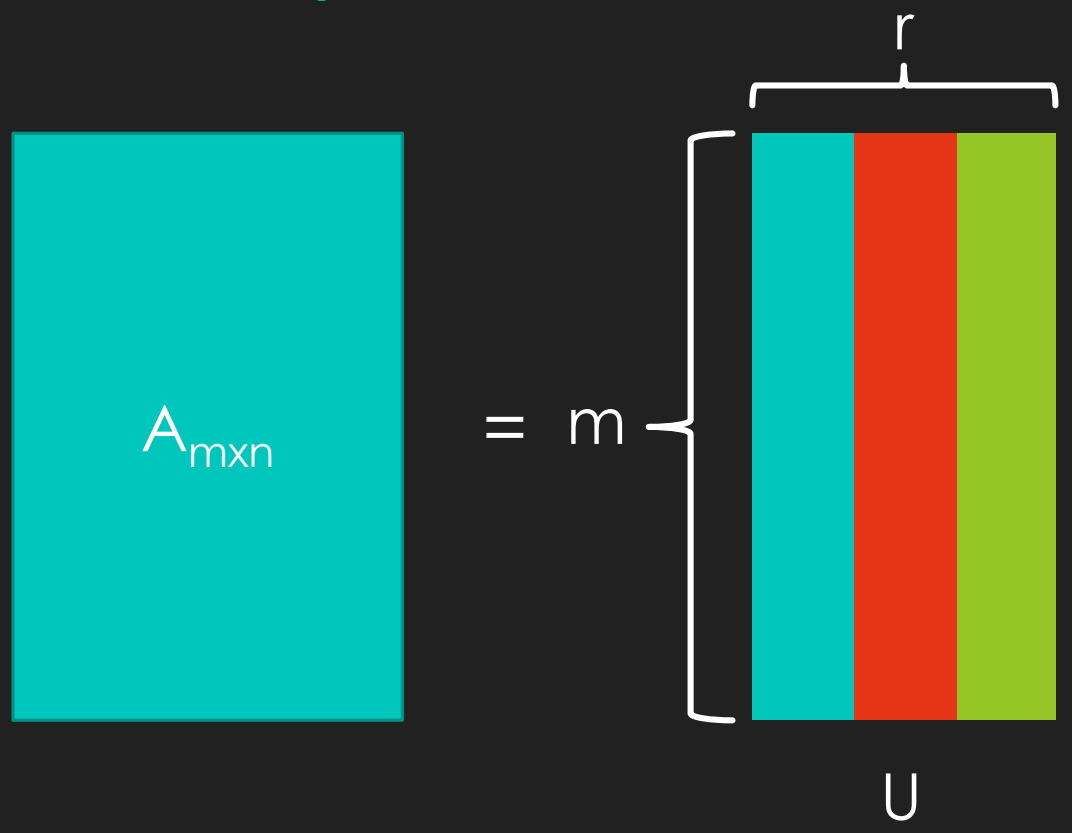
N = Number of columns/words

SVD - Definition

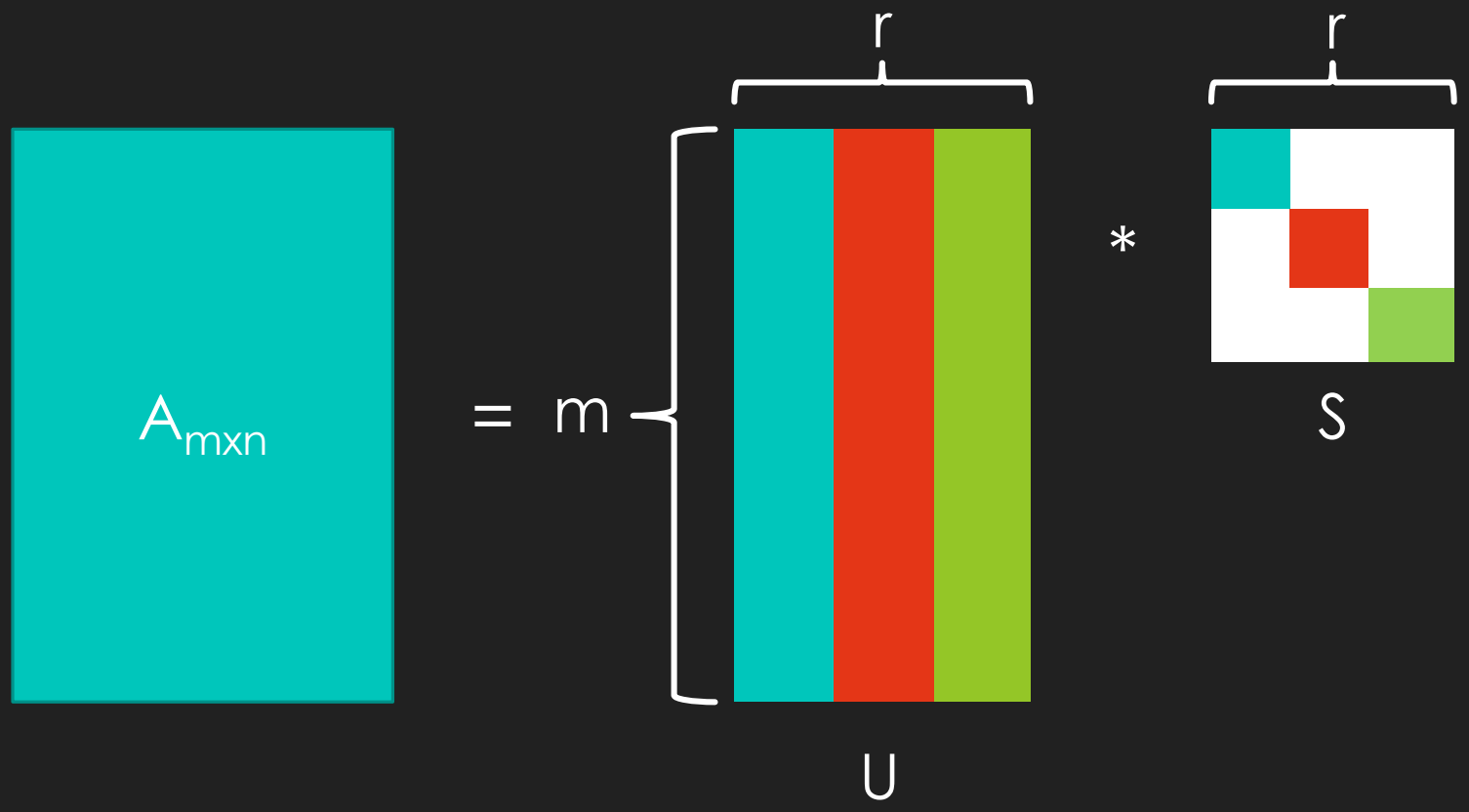
$$A_{[mxn]} = U_{[mxr]} * S_{[rxr]} * (V_{[nxr]})^T$$

- A : Input Data Matrix
  - m x n matrix (m = number of documents, n = number of words/features)
- U : Left Singular matrix
  - m x r matrix (m = number of documents, r = number of concepts)
- S : Rank Matrix
  - r x r matrix (r = rank of A)
- V : Right Singular Matrix
  - n x r matrix (n = number of words/features, r = number of concepts)

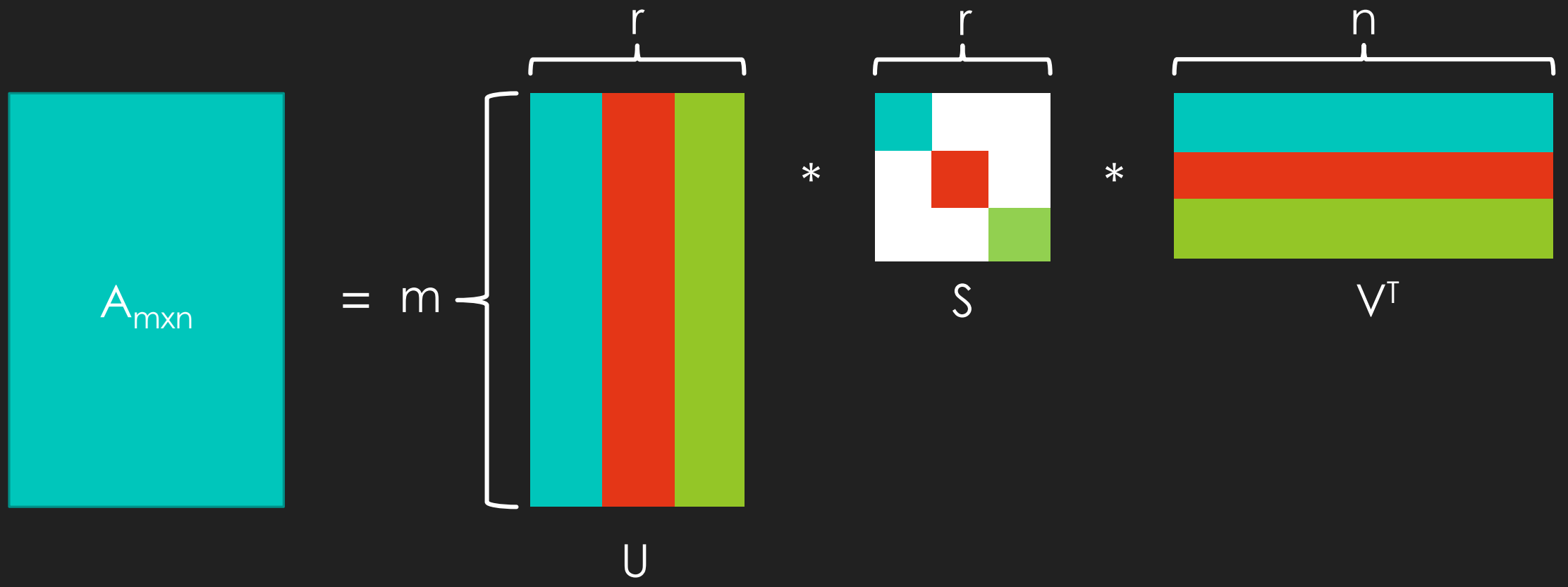
SVD – Visually Explained



SVD – Visually Explained



SVD – Visually Explained



Latent Semantic Analysis – Applications

- Article Bucketing in Websites
- Finding relations between articles/words
- Page Indexing in Search Engines