

Latente Semantische Analyse von Folksonomies

Christoph Wieser
Institut für Informatik

Crowdsourcing, Swarm Intelligence, Data Mining und die Wissenschaften
Workshop an der Ludwig-Maximilians-Universität München

21. Januar 2013

Anfrage-Ergebnis

Volltext: „Tau“

1



Bildersuche nach „Tau“



Tags

Tauwerk

Seil

Tau



Tags

Blatt

Tau

Tropfen



Tags

Mädchen

Seil

Stange

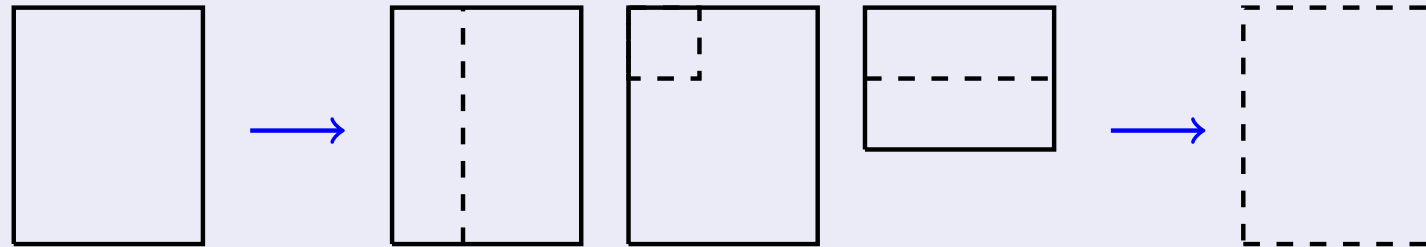
Term-Dokument-Matrix



Tauwerk	1	0	0
Seil	1	0	1
Tau	1	1	0
Blatt	0	1	0
Tropfen	0	1	0
Mädchen	0	0	1
Stange	0	0	1

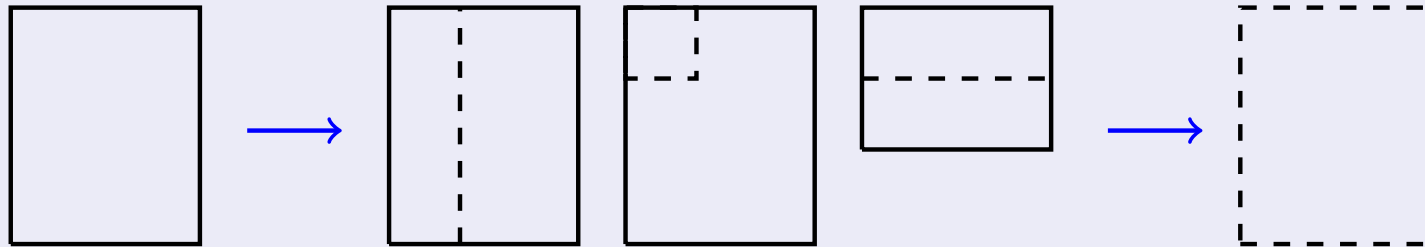
Latent Semantic Indexing und Anfrage (Deerwester, 1990)

Singulärwertzerlegung (SVD) und Dimensionsreduktion



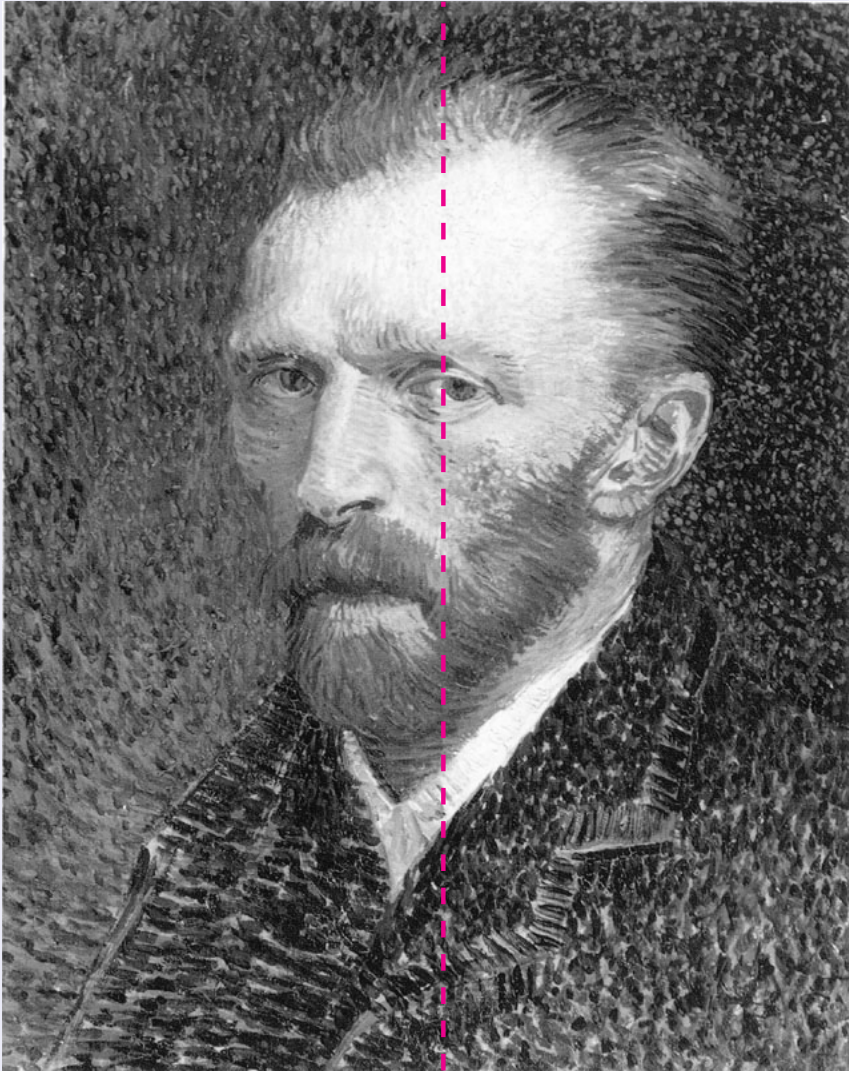
Latent Semantic Indexing und Anfrage (Deerwester, 1990)

Singulärwertzerlegung (SVD) und Dimensionsreduktion

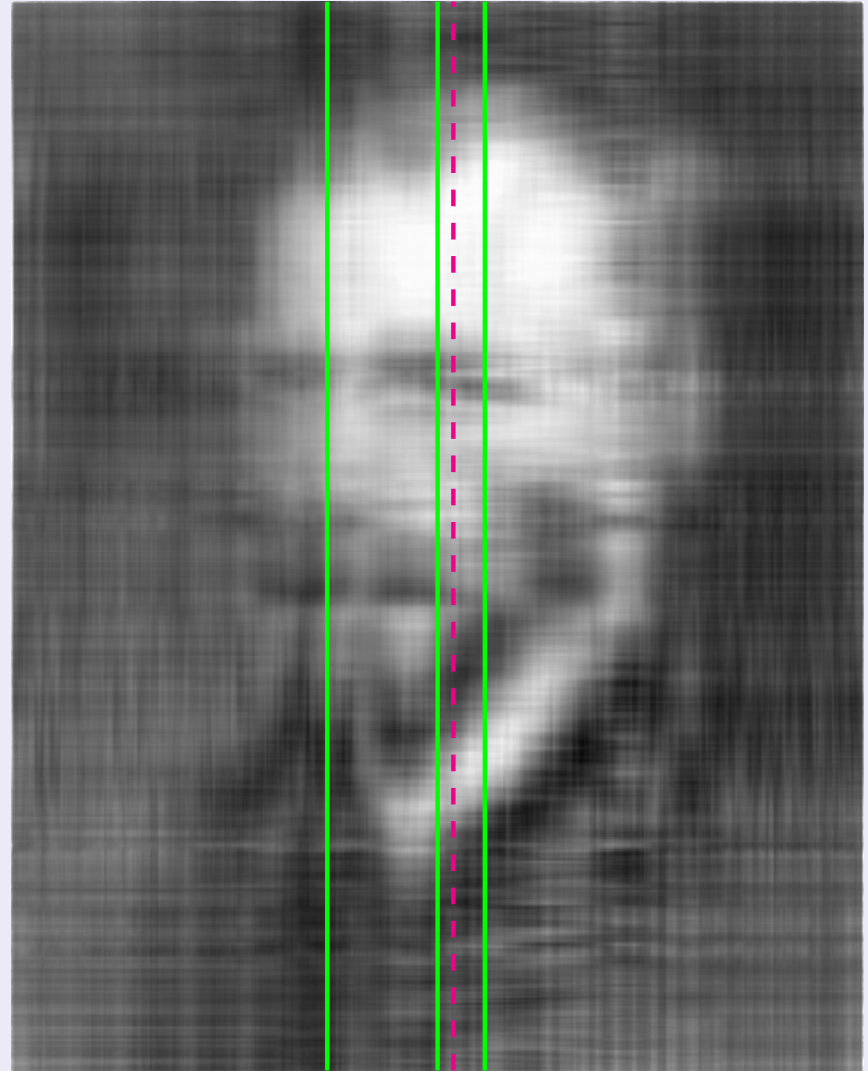


Latent Semantic Indexing als Bild

Term-Dokument-Matrix



Semantischer Raum



Anfrage-Ergebnis

Volltext: „Tau“

1



LSI: „Tau“

1



2



LSI: „Tauwerk“

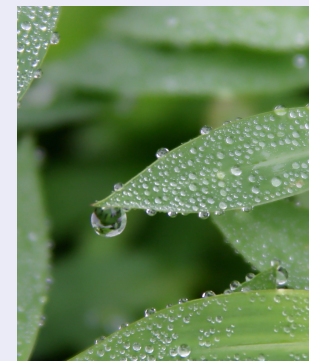
1



2



3



Bildersuche nach „Tau“



Tags

Tauwerk

Seil

Tau



Tags

Blatt

Tau

Tropfen



Tags

Mädchen

Seil

Stange

Bildersuche nach „Tau“ in einer Folksonomy



Tags

Tauwerk

Seil

Tau



Tags

Blatt

Tau

Tropfen



Tags

Mädchen

Seil

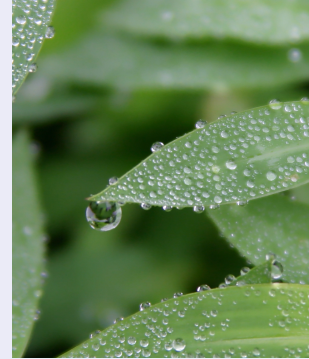
Stange

Term-Dokument-Matrix



Tauwerk	1	0	0
Seil	1	0	1
Tau	1	1	0
Blatt	0	1	0
Tropfen	0	1	0
Mädchen	0	0	1
Stange	0	0	1

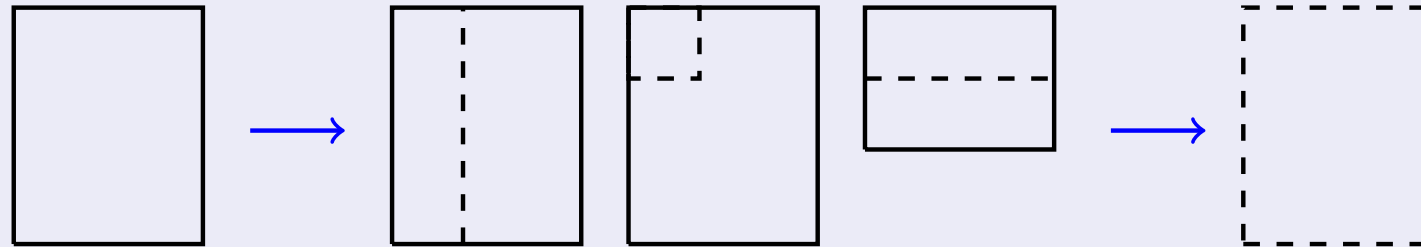
Folksonomy



Tauwerk	1	0	0
Seil	1	0	1
Tau	1	1	0
Blatt	0	1	0
Tropfen	0	1	0
Mädchen	0	0	1
Stange	0	0	1

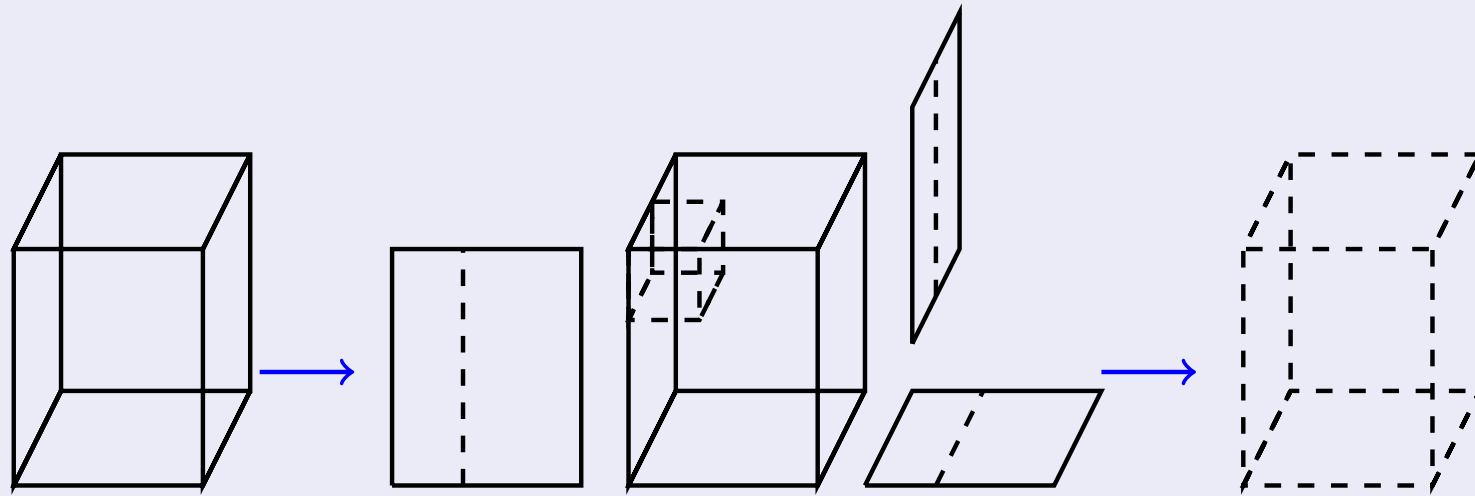
Latent Semantic Indexing und Anfrage (Deerwester, 1990)

Singulärwertzerlegung (SVD) und Dimensionsreduktion



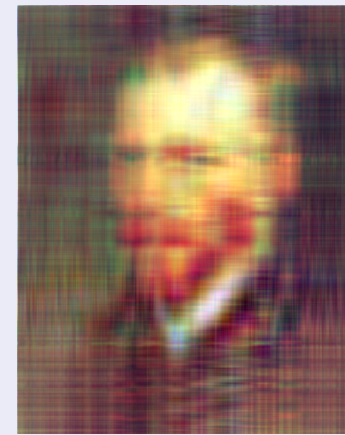
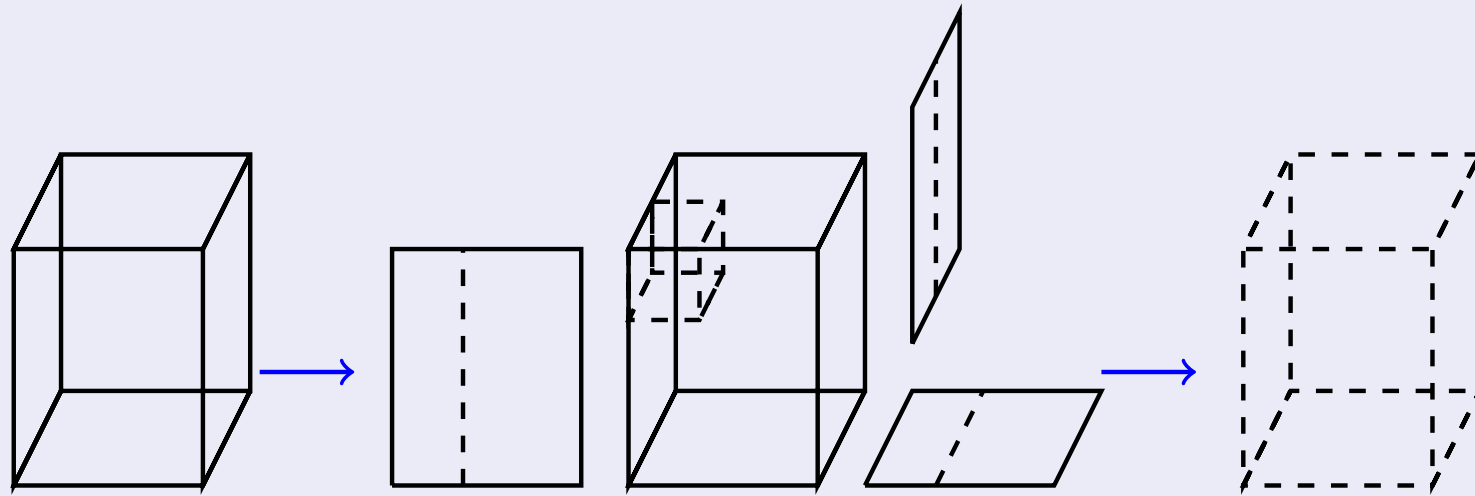
Higher Order Latent Semantic Indexing

Higher Order SVD und Dimensionsreduktionen



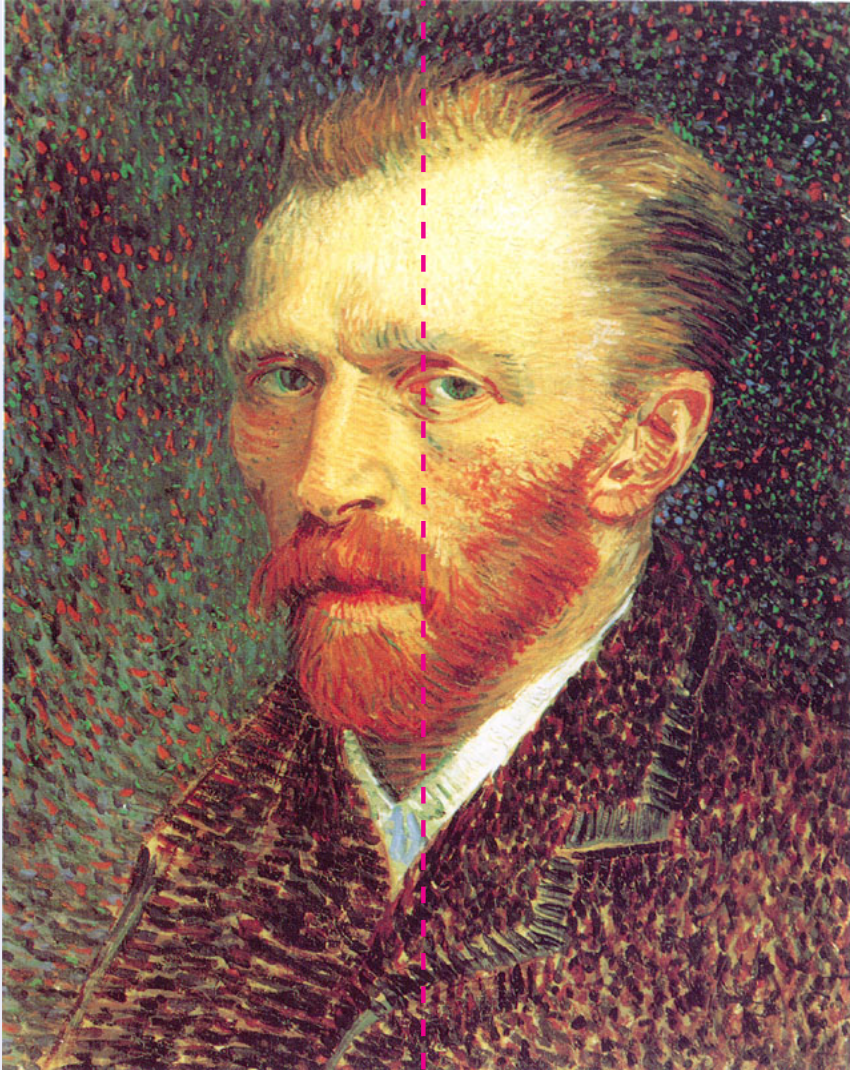
Higher Order Latent Semantic Indexing

Higher Order SVD und Dimensionsreduktionen

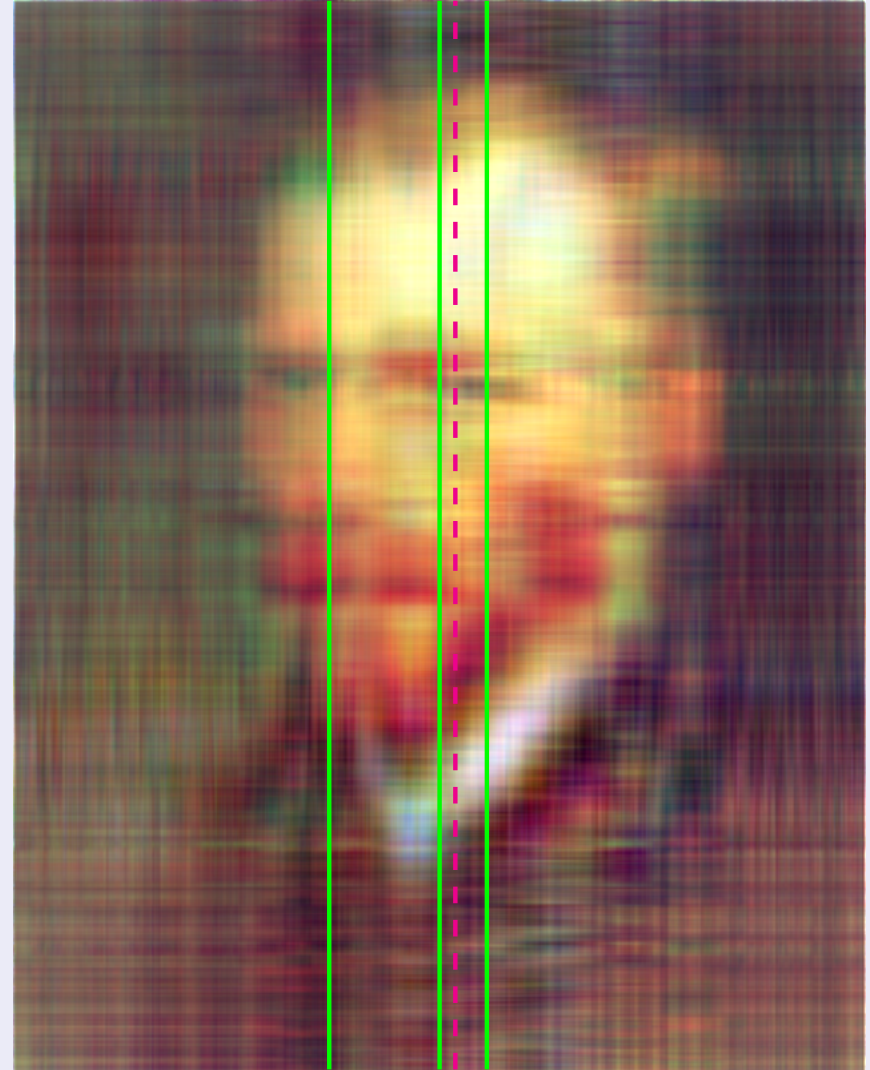


Higher Order Latent Semantic Indexing als Bild

Term-Dokument-Matrix

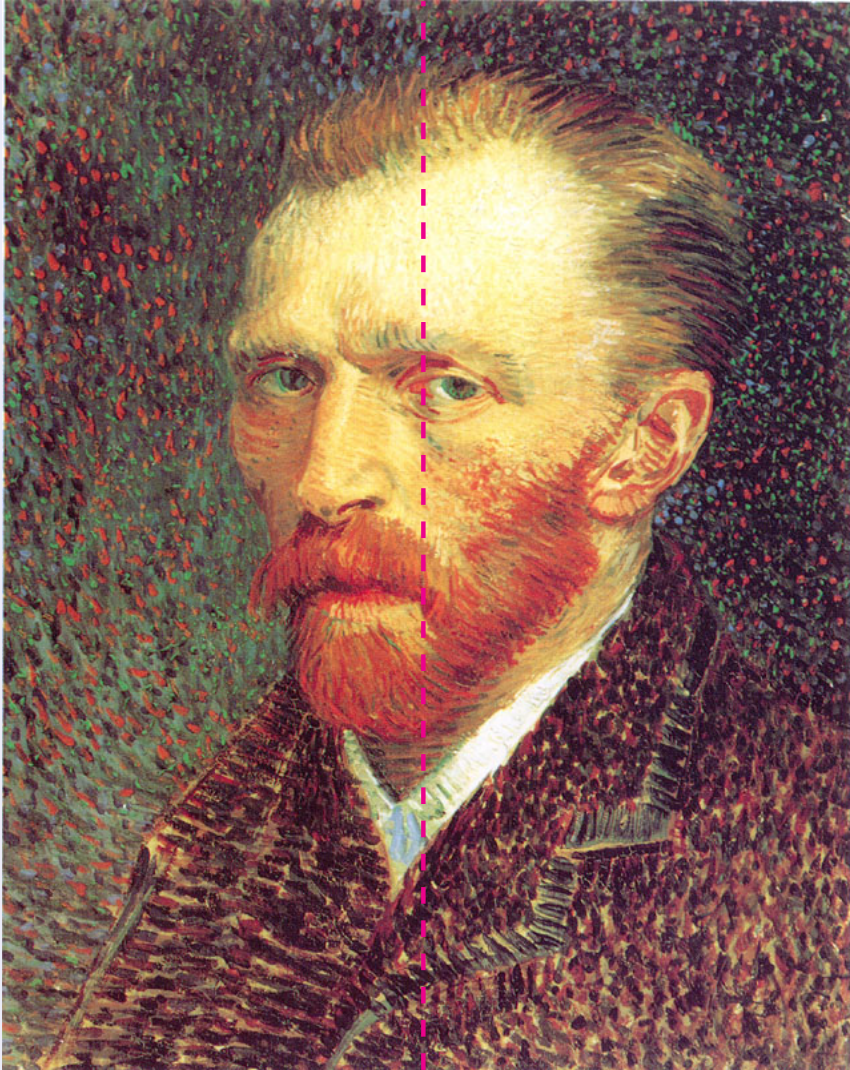


Semantischer Raum

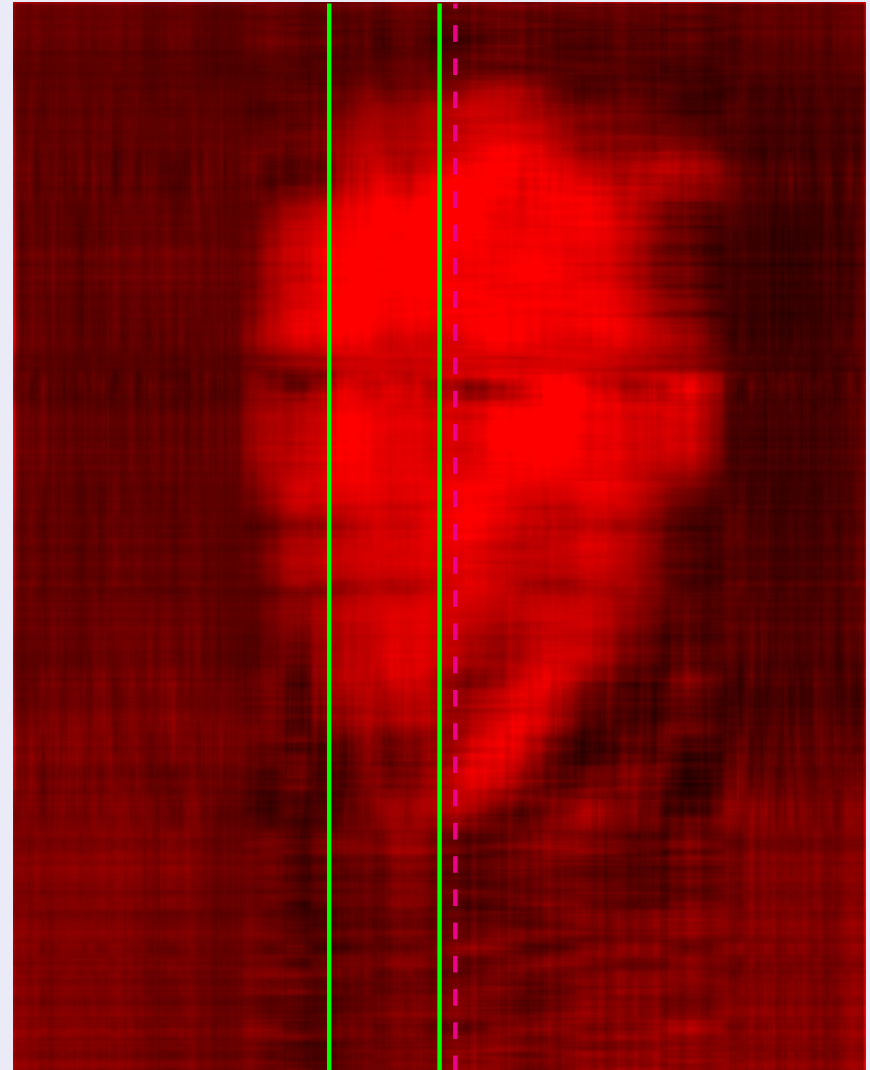


Higher Order Latent Semantic Indexing als Bild

Term-Dokument-Matrix

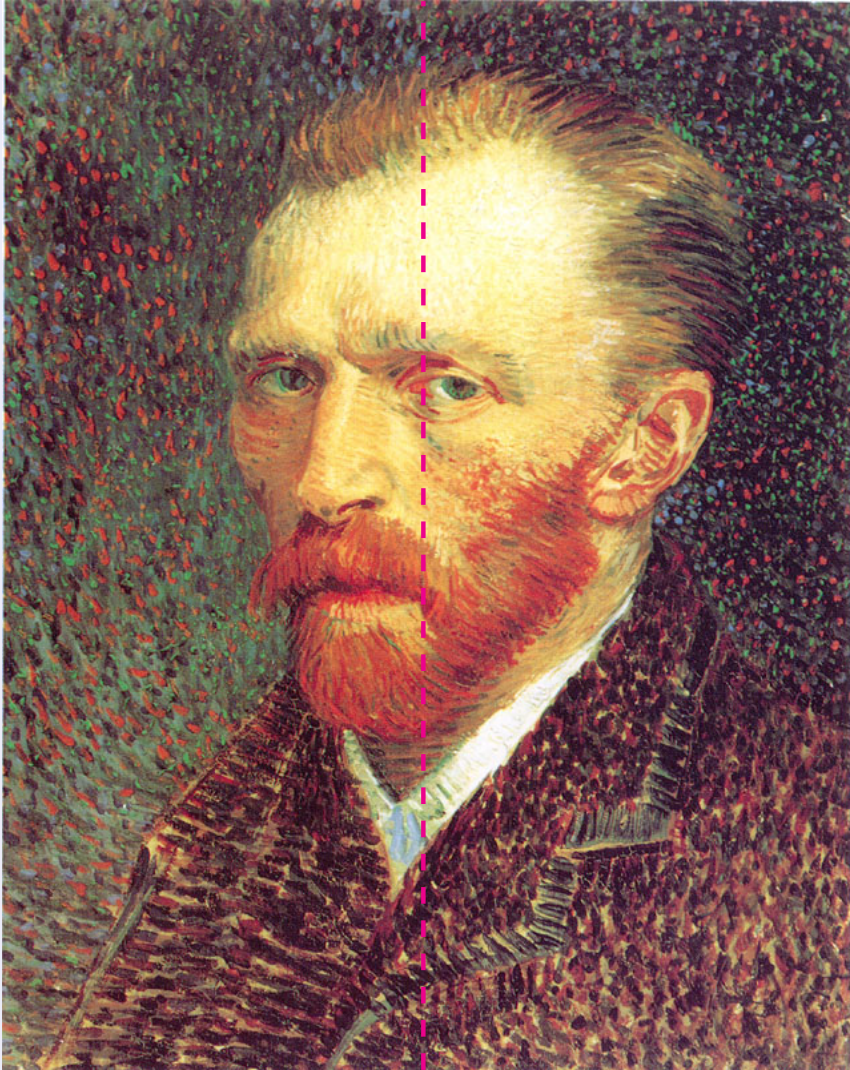


Semantischer Raum **personalisiert**

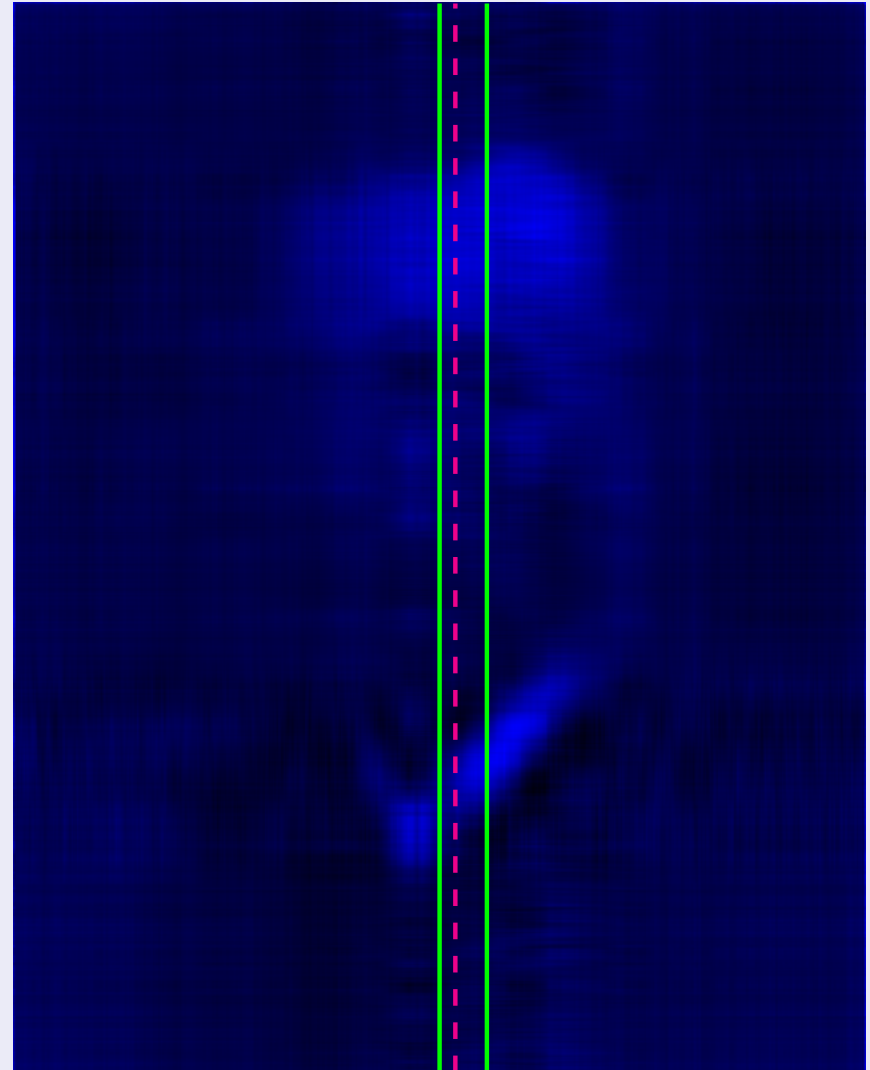


Higher Order Latent Semantic Indexing als Bild

Term-Dokument-Matrix



Semantischer Raum **personalisiert**



Fazit

Verfahren

- Berechnung der Hauptkomponenten (Konzepte) von Folksonomies
- Ähnlichkeitssuche mit Ranking über Kosinusähnlichkeit

Lösungen

- Synonymie-Problem: Tau und Seil
- Homonomie/Polysemie-Problem: Tau als Seil und als Tropfen
- Semnatische Suche: Suche nach Seil liefert Tau
- Personalisierte Suche mit Higher Order LSI

Probleme

- Parameterwahl bei Dimensionreduktion(en)
- Rechenintensive Analyse